UNIVERSITY of CALIFORNIA
Santa Barbara

Topics in Sequential Decision Making: Analysis and Applications

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Mechanical Engineering

by

Sandra Hala Dandach

Committee in charge:

Professor Francesco Bullo, Chair
Professor João Hespanha
Professor Bassam Bamieh
Professor Jeff Moehlis

June 2011

The dissertation of Sandra Hala Dandach is approved:

Professor Jeff Moehlis

Professor Bassam Bamieh

Professor João Hespanha

Professor Francesco Bullo, Chair

June 2011

Topics in Sequential Decision Making: Analysis and Applications

by

Sandra Hala Dandach

To my parents Sabat Daher and Rakan Dandach, who taught me everything I know about morals, passion, courage, dedication and patriotism. Your love and support, was and will continue to be my biggest inspiration and source of strength.

Acknowledgements

I would like to thank Francesco Bullo for all the lessons he taught me and for all the patience he had with me. I learned from him how important it is to take care of the details, no matter how early a stage the problem is in. I also learned the importance of mathematical correctness, and how writing things in details helps you evaluate your understanding of the problem, and take things in the right direction. Working with him was a pleasure, and if I take nothing else from him but his enthusiasm for science, that would be a gift for which I will thank him for years and years to come. Finally, Francesco's support and kindness were overwhelming. I cannot think of a better role model and mentor to have in life, and that statement is not restricted to my professional life.

I was very lucky to collaborate with a great researcher and a wonderful friend. Ruggero Carli's input and collaborations to this thesis were great. I learned a lot working with Ruggero about the importance of proper problem formulations and the greater importance about the necessity of reiterating on failed solution attempts. I was even luckier to have him as a friend outside the lab. I look forward for long years of collaboration.

I would like to thank Bassam Bamieh who had various inputs and suggestions for my research. Sometimes he would suggest an approach, or answer a questions, other times he would suggest a way of presenting my results and very often he would suggest interesting directions. I would additionally like to thank Bassam for taking the time to do all that and for the friendship we developed. Joao Hespanha was a great teacher and a great committee member. He has this talent of understanding things to the detail that keeps on amazing me, his questions really make you think! and his clarity of mind is inspiring. Jeff Moehlis was always available and always helped me clarify the message I am trying to pass. I would like thank Joao and Jeff for all our interactions; they were very educational to me.

I would like to thank Petar Kokotovic who is without any doubt the godfather of anyone who has the chance of studying control theory in CCDC at UCSB. He was always available and will always find interest in any problem you come to him with. I would also like to thank PK for being a live example of how it is possible to "walk with kings- nor lose your common touch". He touched many lives I am sure, but I can definitely confirm he touched mine. Mustafa Khammash was an adviser, became a friend, then again became a mentor. I had the chance of interacting with him on a large variety of topics and I look forward for a lot of interactions in the future. I would like to thank him for his great friendship and for one specific long and very nice conversation we had that I will never forget. It taught me a lot about myself. Thank you Mustafa for yet another lesson.

Before I came to UCSB I had the chance of being trained on the hands of great scientists. I thank Soura Dasgupta who taught me from scratch about research techniques. Fuad Mrad was my first mentor and he taught me to believe in myself and to do what I believe in regardless of the noise around me. For that alone, I owe him more than I can ever give back.

Motion lab and its members constitute a big part of my graduate life. I owe them all thanks for many long fun nights spent in the lab and many exciting discussions that always ended up on the white board that can never be erased...

My friends in Santa Barbara had a major role in making this place a home away from home. I want to thank Chris Hammetter, Danielle Sedillo, Bryan Eisenhower, Scott Ferguson, Kelly Schmandt, Miguel Jimenez, Laura Oropoza, Steve Smith, Marco Zuliani, Alessandra Sala, Anahita MirTabatabaei, Joey Durham, Hussam Mousa, Payam Bozorgi, Mohammad Nasr Azadani, Ruggero Carli for the unforgettable years.

Many thanks are due to friends whose friendship survived the longest of distances Youssef Brarake, Najwa Hamzeh, Joy Jadam, Jad Bejjani. They were supportive until the end. I could not have done much wihtout such support.

Saadeh Abou Nassif, your departure continues to be a wound that defies time. Thank you for the great inspiration that you were and will always be.

I have acquired an additional sister in Santa Barbara: Ceren Budak. She has known me through the ups and downs and was always a true sister. I thank her and UCSB family housing for that.

Finally, I want to thank my family who never stopped feeding me with their love and support through it all. I know for a fact, I would have not been here if it were not for you. Sabat Daher, Rakan Dandach, Diala Dandach, Patricia Dandache and Mouin Abdallah, Moustapha Dandache and Jad Dandach please accept my thanks for all the support you showed me in my life.

Curriculum Vitæ

Sandra Hala Dandach

Education

Ph.D. University of California, Santa Barbara, CA, USA.
Mechanical Engineering. Sep'07- Jun'11.
Center for Control, Dynamical Systems and Computation.
Advisor: Francesco Bullo.
Thesis: Distributed decision making.

M.Sc. University of California, Santa Barbara, CA, USA.
Mechanical Engineering. Sep'05-Sep'07.
Center for Control, Dynamical Systems and Computation.
Advisor: Mustafa Khammash.
Thesis: Bacterial competence: stochastic strategies for survival.

M.Sc. University of Iowa, Iowa city, IA, USA.
Electrical Engineering. Sep'03-Aug'05.
Advisor: Soura Dasguta.
Thesis: Stability of adaptive delta modulation with and without a forgetting factor.

B.E. American University of Beirut, Beirut, Lebanon.
Electrical Engineering. Sep'98-Jul'02.
Area of emphasis: Automation, robotics and control.

Research Interests

1. Decision theory, detection theory, estimation theory, decision making.
2. Systems biology, computational methods.
3. Optimal space allocation, distributed algorithms design, control theory, optimization theory.
4. Control over communication channels, signal processing in control.

Technical Skills

- **Languages:** Fluently speak, read and write: English, French and Arabic.
- **Computer skills:** Mathematica, Matlab, Labview, PLC programming, C++, basic knowledge of C.

- **Laboratory skills:** basic knowledge on how to perform biochemical experiments (CheY) cloning, protein purification, PCR, agarose gel electrophoresis and protein data analysis.
- **Personal skills:** Very active, very fast learner, great team player, great presentation skills.

Technical Reviewer Automatica, Journal of Guidance, Control and Dynamics, IEEE Conference on Decision and Control, IFAC World Congress, American Control Conference, European Control Conference, IFAC Symposium on System Identification.

Professional Affiliations Biophysical Society, Institute of Electrical and Electronics Engineers (IEEE), IEEE Control Systems Society (IEEE CSS), IEEE Computer Society, IEEE Women in Engineering, IEEE Aerospace and Electronic Systems Society, IEEE Robotics and Automation Society.

Research Experience

Graduate Student Researcher, University of California, Santa Barbara. Sep'07-present. Investigated two different aspects of decision theory on a network:

1. Cooperative detection. Major contributions:
   - Designed distributed algorithms to solve regional localization as a multi-hypothesis problem.
   - Provided convergence results and performance measures of the suggested algorithms.
   - Improved the performance of the algorithms by optimally partitioning the environment.
2. Cooperative decision making. Major contributions:
   - Provided a novel computational method that allowed exact analysis of the accuracy and time of a network of cooperative agents.
   - Conducted sensitivity analysis for two special rules, and showed that the performance and decision time for large networks are defined by the performance at special times for a single individual.
   - Showed that the optimal fusion rule varies with the local fusion rules and the network size as well as the desired performance.

Graduate Student Researcher, University of California, Santa Barbara.Sep'05-Aug'07. Studied the problem of stochastic switching in Bacillus-subtilis bacteria. Major contributions:

1. Provided a novel computational method that gave analytical answers to the Chemical Master Equation often used in the analysis of genetic circuits.
2. Investigated the expected time and the sensitivity of the stochastic switching in cells to various binding affinities, transcription rates, degradation rates, etc.

Graduate Student Researcher, University of Iowa, Iowa City. Sep'03-Aug'05.

Investigated the following two problems:

1. Stability of adaptive delta modulation (ADM). Major contributions:
   - Proved that the classical ADM algorithm enters in periodic cycles.
   - Modified the ADM algorithm to limit the magnitude of the oscillations.
2. Source localization with mobile agents. Major contributions:
   - Designed a continuous-time algorithm that localizes a source without amplification of noise.
   - Showed that the algorithm converges exponentially fast under mild conditions on the agent motion and continues to successfully localize a slowly moving source.

Teaching Experience

Teaching Assistant, University of California, Santa Barbara. Sep'05-present.
Courses: Control Systems Design I, Control System Design II, Engineering Mechanics: Dynamics.
Duties included teaching weekly discussion sections, teaching several classes, preparing laboratory experiments, leading laboratory sessions, grading assignments as well as midterms and finals, assigning tasks to other teaching assistants, ensuring coordination between the various teaching assistants.
Teaching Assistant, American University of Beirut, Lebanon. Sep'02-Sep'03

Course: Instrumentation, sensors and actuators.

Duties included help leading weekly laboratory sessions, grading assignments.

Publications

Journal Articles

1. S.H. Dandach, R. Carli and F. Bullo, **Sequential Decision Aggregation: Accuracy and Decision Time for Decentralized SPRT,** *Proceedings of the IEEE*, in *The special issue on Interaction Dynamics at the Interface of Humans and Smart Machines*. (Accepted. Available at http://arxiv.org/abs/1004.0356)
2. S.H. Dandach and F. Bullo, **Distributed Sequential Algorithms for Regional Source Localization**, *Automatica*. (Submitted. Available at http://motion.mee.ucsb.edu/~sandra/Seq_loc.pdf)
3. S.H. Dandach and M. Khammash, **Analysis of stochastic strategies in bacterial competence: a master equation approach**, *PLoS Computational Biology*, 6(11), 2011.
4. S.H. Dandach, B. Fidan, S. Dasgupta and B.D.O. Anderson, **A Continuous Time Linear Adaptive Source Localization Algorithm, Robust to Persistent drift**, *Systems and Control Letters*, 58(1):7-16,2009.
5. S.H. Dandach, S. Dasgupta and B.D.O. Anderson, **Stability of Adaptive Delta Modulators with Forgetting Factor and Constant Inputs**, *International Journal of Adaptive Control and Signal Processing*. (Accepted. Available at http://motion.mee.ucsb.edu/~sandra/acsadm_revision.pdf)

Conference Publications

1. S.H. Dandach and Mustafa Khammash, **A Novel Computational Method for Stochastic Strategies for Bacterial Survival Analysis**. Invited session on "Dynamics and Control of Cellular Systems", in *American Control Conference*, 2011, San Francisco, CA. (Accepted)
2. S.H. Dandach, R. Carli and F. Bullo, **Accuracy and Decision Time for a Class of Sequential Decision Aggregation Rules**. *Conference for Decision and Control*, 2010, Atlanta, GA.
3. S.H. Dandach, R. Carli and F. Bullo, **Accruacy and decision time for cooperative implementations of the sequential probability ratio test**. Invited session on "humans-in-loop systems", in *American Control Conference*, Baltimore, MD, June 2010.
4. S.H. Dandach and F. Bullo, **Algorithms for regional source localization**, in *American Control Conference*, St. Louis, MO, pages 5440-5445, June 2009.
5. S.H. Dandach, B. Fidan, S. Dasgupta and B.D.O. Anderson, **Adaptive source localizations by mobile agents**, in *IEEE Conference on Decision and Control*, pages 2045-2050, San Diego, CA, December 2006.

6. S.H. Dandach, S. Dasgupta and B.D.O. Anderson, **Stability of adaptive delta modulators with a forgetting factor and constant inputs**, in *IEEE Conference on Decision and Control and the European Control Conference*, pages 5808-5813, Seville, Spain, December 2005.
7. S.H. Dandach, S. Dasgupta and B.D.O. Anderson, **Stability of adaptive delta modulators with constant inputs**, in *IASTED International Conference of Networks and Communication Systems*, Krabi, April 2005.
8. S.H. Dandach, S. Dasgupta and J. Freudenberg, **Control over bandlimited communication in channels: Intersampling performance**, in *Proceedings of the International Conference on Systems, Man and Cybernetics*, pages 3886-3601. The Hagues, October 2004.
9. S.H. Dandach, S. Dasgupta, **Optimal design of stable haptic interfaces**, in *Proceedings of SICE 2004 Annual Conference*, Sapporo, Japan, August 2004.
10. F.Mrad, S.H. Dandach, S. Azar and G. Deeb, **Operator-friendly common sense controller with experimental verification using LabVIEW**, in *Proceedings of the 2005 International Symposium on Intelligent Control*, Cyprus, June 2005.

Abstract

Topics in Sequential Decision Making: Analysis and Applications

by

Sandra Hala Dandach

Interest in group decision making spans a wide variety of domains. Be it in electoral votes in politics, Bayesian learning in social networks, distributed detection in robotic and sensor networks, or cognitive data processing in the human brain, establishing the best strategy or understanding the motivation behind an observed strategy, has been of interest for many researchers. This thesis studies two sequential decision making problems, in the first problem the individuals do not communicate with each other, while in the second problem they are allowed to exchange information.

In the non-cooperative setting, we consider a collection of agents, each performing binary hypothesis testing and obtaining a decision over time. We assume that the agents are identical and receive independent information. Individual decisions are sequentially aggregated via a threshold-based rule. In other words, a collective decision is taken as soon as a specified number of agents report a con-

cordant decision (simultaneous discordant decisions and no-decision outcomes are also handled). We relate the accuracy and decision time of the whole population, to the accuracy and decision time of a single individual and to the fusion rule. We also provide a scalability analysis for some group decision rules and show that in the limit of large group sizes, the accuracy and decision time of the group are dictated by the accuracy and decision time of a single individual.

In the cooperative setting, a group of individuals are monitoring an environment and answering a question about the location of a source. The environment is divided into smaller regions of responsibilities, each individual is responsible for one or multiple regions. We pose the problem as a multiple hypothesis testing problem and design a distributed sequential localization algorithm with guaranteed accuracy bounds; we also provide a proof of almost sure convergence of our algorithm in the limit of a large numbers of measurements. We pose and distributedly solve optimization problems whose solution provides a choice of regions that improves the performance of the localization algorithm. We illustrate the applicability of the proposed distributed optimization algorithm to a family of optimization problems.

Contents

List of Figures

Chapter 1

Introduction

Decision making happens so often and in such various contexts that it is slightly difficult to relate it to a single science. How and why we do what we do, can be affected by who we are and how we interact with our environment. The effect might come, but is not limited to, people around us, our personalities and prejudgment, our personal preferences, etc. Recently, this topic has captured the interests of many scientists, each of which concentrated on understanding different parts of the problem. Some researchers proposed models that study the effect of rewards and feedback on individuals shedding light on motivation and character-related features that are specific to each individual, while other researchers were interested in studying the way decisions are made in the brain on a more detailed level from the dynamics perspective. The work presented in this thesis falls somewhere between the two families of problems. In this work, we analyze how information processing affects the overall group decision, which sets the ground to understanding questions about group information processing (e.g. cognitive information processing in a group of neurons) if information is assumed to be available about the dynamics of the individuals (e.g. decision making at each neuron level). We note that an additional motivation of the work presented in this thesis is the design of distributed algorithms in cooperative robotic networks, where understanding the effect of the interaction between the robots as well as understanding the decision making of each stand-alone robot is useful in designing proper algorithms to perform a task of interest. We start by presenting the various problems studied in this thesis.

Accuracy and Decision Time for Sequential Decision Aggregation Work in this thesis related to *sequential decision aggregation* aims to understand how sequential processing of decisions from sequential decision makers affects the speed and accuracy with which these individuals reach a collective decision. This class of problems has a rich history and some of its variations are studied in the context of distributed detection in sensor networks and Bayesian learning in social networks.

In our problem, a group of individuals independently decide between two alternative hypotheses, and each individual sends its local decision to a fusion center.

The fusion center decides for the whole group as soon as one hypothesis gets a number of votes that crosses a pre-determined threshold. We are interested in relating the accuracy and decision time of the whole population, to the accuracy and decision time of a single individual.

Distributed Sequential Algorithms for Regional Source Localization Applications where source localization is of great concern vary between finding the source of oil spills in the ocean, determining cellular locations, detecting an earthquake's epicenter, locating an acoustic source, or simply finding an intruder in a protected environment. For most of these applications, it is sufficient to find a region that contains the source rather than pinpointing the exact source position, which relies most of the time on approximations.

In this thesis, we consider the following problem: A source at an unknown location in a bounded region Q transmits a power signal. N sensors receive noisy and decayed versions of the signal, they can communicate and exchange measurements. The environment Q is divided into M regions W_α, where $\alpha \in \{1, \dots, M\}$. The objective of the sensors is to find which region contains the source.

Gossip algorithm for a class of environment partitioning problems with separable rewards and equitability constraints Optimization problems where a given region is divided into N sub-regions have applications in various fields. We name, among others, districting, facility design, warehouse layout, urban planning, etc. In each of these problems the objective function has certain characteristics which are usually useful in tailoring a proper solution. In some problems, the cost function is such that the optimality of partitions can be specifically characterized, for example for some problems Voronoi partitions or other general forms of Voronoi partitions were proved to be optimal. For more general problems, such characterization is not possible.

In this thesis we study this group of problems and present gossip based algorithms that solve environment partitioning problems where one part of the problem is an equitability constraint, and where the objective function depends on each variable independently from the other.

1.1 Literature review

In this section we give a brief literature review of the main references of the various topics or tools mentioned in this thesis.

1.1.1 Hypothesis testing and decision making

The framework we analyze in this Chapter 2 is related to the one considered in many papers in the literature, see for instance [1, 2] and references therein.

The focus of these works is mainly two-fold. First, researchers in the fields aim to determine which type of information the decision makers should send to the fusion center. Second, many of the studies concentrate on computing optimal decision rules both for the individual decision makers [3] and the fusion center, where optimality refers to maximizing accuracy. One key implicit assumption made in numerous works is that the aggregation rule is applied by the fusion center only after all the decision makers have provided their local decisions.

Tsitsiklis in [4] studied the Bayesian decision problem with a fusion center and showed that for large groups identical local decision rules are asymptotically optimal. Varshney in [5] proved that when the fusion rules at the individuals level are non-identical, threshold rules are the optimal rules at the individual level. Additionally, Varshney proved that setting optimal thresholds for a class of fusion rules, where a decision is made as soon as a certain number q out of the N group members decide, requires solving a number of equations that grows exponentially with the group size. The fusion rules that we study in this work fall under the q out of N class of decision rules. Finally, Varshney proved that this class of decision rules is optimal for identical local decisions.

One major difference between these fusion rules and previously analyzed similar rules [6, 7, 8, 9], is the notion of time. While in the literature, researchers have already studied the accuracy of a group aggregating decisions from various

DMs under different fusion rules, one common assumption was implicitly made by all the work we came across. This assumption is that the group applies the fusion rule only after all individuals have decided. For more details see [10] and references therein.

Researchers in behavioral studies refer to the decision making scheme where everyone is given an arbitrary the time to respond as the free response paradigm. Since the speed of the group's decision is one of our main concerns, we found it necessary to adjust the analysis in a way that makes it possible to compute the joint probabilities of each decision at each time instant. For more detailed coverage of this topic we refer the reader to [11, 12, 13] and references therein. Many researchers proposed mathematical models that aimed to explain observations made in the cognitive science literature. Among these works we mention [14, 15, 16, 17].

The multiple hypothesis problems are considerably more difficult than the binary problem and optimality of the proposed algorithms is usually hard to prove. Some tests that have some asymptotic optimality properties were developed in the literature, but these tests tend to be very complex [18, 19, 20]. Alternatively ad hoc tests based on repeated pairwise applications of optimal sequential hypothesis tests [3] were developed but these tests have little optimality results, e.g., see [21].

1.1.2 Source localization

In the classical source localization problem, a number of sensors collaborate to locate the exact position of a source. The relation between the position of a source and the received signal strength (RSS) is described in [22, 23]. A survey of the literature is presented in [24]. Several authors treat localization as a nonconvex optimization problem [25, 26]. Gradient descent algorithms and weighted least squares approximations can be used to solve the maximum likelihood estimation problems but such algorithms tend to get stuck at local optima [27, 28]. Authors in [29] approximate the nonlinear nonconvex optimization problem by a linear and convex problem. Hero et al. in [30] use a method of projection onto convex sets. A necessary and sufficient condition for the convergence of this algorithm is that the source lies inside the convex hull of the sensors. Properly placing the sensors assumes knowledge of the position of the source.

1.1.3 Distributed algorithms and environment partitioning

Designing distributed algorithms is in general a problem specific task, and many researchers from various communities have looked at this problem, see [31, 32, 33, 34, 35] and references therein. One of the many interesting problems studied over networks is average consensus. We will not cover this topic in details here, but we will list the work presented by Boyd et al [36] that introduces the distributed gossip algorithm to solve the consensus problem. In their work, Boyd et al. prove that the gossip algorithm will achieve consensus and reach the average of the initial conditions of all the states of the nodes in a network. As we will show in the sequel, the gossip algorithm is a natural fit to the distributed regional optimization problem we look to solve.

In addition to optimality, sometimes a notion of equal load or measure is required, and the problem becomes one of finding equitable partitions. Many recent papers study the problem of equitable partitioning see for instance [37, 38, 39, 40, 41] and references therein.

Optimally partitioning an environment into regions that improve an objective function is yet another problem of interest. The isoperimetric problem is an example of geometric problems where a geometric shape is sought as to maximize a function (isoperimetric ratio) while maintaining a given measure (the area covered) . Mathematicians had conjectured for centuries that the circle is the one geometric figure that satisfies this property. More complicated problems, with closer connections to partitioning problems, had been proposed among which the one we study in this chapter. Our objective is to cover an environment without leaving empty spaces while minimizing the sum of the perimeters of all the partitions. The solution to this geometric optimization problem turns out to be partitions obtained by hexagonal tiling. Interestingly, nature seems to foster the first geometer to solve this problem. When building their honeycombs, bees revert to hexagonal tiles. This minimizes the lost effort they need to put producing wax, by minimizing the perimeter of the partitions, while avoiding any empty spaces between the honey combs so that no parasites can grow in their honeycombs [42, 43]. Triangles, squares and hexagons are the only regular polygons that tile the plane. Among them, hexagonal tiles minimize the sum of perimeters of the regions for a given area.

1.2 Contributions of the thesis

Chapter 2 studies prototypical strategies to sequentially aggregate independent decisions. We consider a collection of agents, each performing binary hypothesis testing and each obtaining a decision over time. We assume the agents are identical and receive independent information. Individual decisions are sequentially aggregated via a threshold-based rule. In other words, a collective decision is taken as soon as a specified number of agents report a concordant decision (simultaneous discordant decisions and no-decision outcomes are also handled).

We obtain the following results. First, we characterize the probabilities of correct and wrong decisions as a function of time, group size and decision threshold. The computational requirements of our approach are linear in the group size. Second, we consider the so-called fastest and majority rules, corresponding to specific decision thresholds. For these rules, we provide a comprehensive scalability analysis of both accuracy and decision time. In the limit of large group sizes, we show that the decision time for the fastest rule converges to the earliest possible individual time, and that the decision accuracy for the majority rule shows an exponential improvement over the individual accuracy. Additionally, via a theoretical and numerical analysis, we characterize various speed/accuracy tradeoffs. Finally, we relate our results to some recent observations reported in the cognitive information processing literature.

Chapter 3 studies the problem of source localization as a multiple hypothesis testing problem, where each hypothesis corresponds to the event that the source belongs to a particular region. We use sequential hypothesis tests based on posterior computations to solve for the correct hypothesis. Measurements corrupted with noise are used to calculate conditional posteriors. We prove that the regional localization problem has geometric properties that allow correct detection almost surely in the limit of a large number of measurements. We present the *Sense, Transmit & Test* distributed algorithm that allows sequential sensing, communication and testing and we analyze the accuracy of this distributed algorithm and show that the test ends in a finite time. We also present numerical results illustrating properties of the suggested algorithm.

Chapter 4 studies distributed gossip algorithms that solve a family of environment partitioning problems over a network. This work was inspired by a quest to choose environment partitions that optimize the performance of a localization algorithm. The problem falls under a family of problems in facility design where the objective function depends separately on each optimization parameter. We design a gossip based algorithm that solves a sequence of optimization problems, where the optimization variable are the partitions of the environment. We study four problems of interest. The first problem is the equitable partitioning problem, the second problem is the doubly equitable partitioning problem, the third problem is the isoperimetric problem (and some variation of it) and the fourth problem is the regional localization problem. We present algorithms that solve these problems and present simulation results showing the optimal partitioning to all these problems.

Chapter 5 presents summaries of the results and suggested future research directions.

Chapter 2

Accuracy and Decision Time for Sequential Decision Aggregation

2.1 Introduction

In this chapter we study a group of individual sequential decision makers, where each individual is running its independent binary sequential hypothesis test. Individuals communicate their decisions, as soon as the latter are made, to a fusion center that aggregates the information sequentially, until the number of votes in favor of a hypothesis crosses a threshold. We study the group decision time and accuracy and relate that to the group size and the local accuracies and decision time at the individual level.

2.1.1 Problem description and motivation

We study the following problem: a group of individuals independently decides between two alternative hypothesis, and each individual sends its local decision to a fusion center. The fusion center decides for the whole group as soon as one hypothesis gets a number of votes that crosses a pre-determined threshold. We are interested in relating the accuracy and decision time of the whole population, to the accuracy and decision time of a single individual. We assume that all individuals are independent and identical. That is, we assume that they gather information corrupted by i.i.d. noise and that the same statistical test is used by each individual in the population. The setup of similar problems studied in the literature usually assumes that all individual decisions need to be available to the fusion center before the latter can reach a final decision. The work presented here relaxes this assumption and the fusion center might provide the global decision much earlier than the all individuals in the group. Researchers in behavioral studies refer to decision making schemes where everyone is given an equal amount of time to respond as the "free response paradigm." Since the speed of the group's decision is one of our main concerns, we adjust the analysis in a way that makes it possible to compute the joint probabilities of each decision at each time instant.

2.1.2 Chapter contributions

The contributions of this chapter are three-fold. First, we introduce a recursive approach to characterize the probabilities of correct and wrong decisions for a group of sequential decision makers (SDMs). These probabilities are computed as a function of time, group size and decision threshold. The key idea is to relate the decision probability for a group of size N at each time t to the decision probability of an individual SDM up to that time t, in a recursive manner. Our proposed method has many advantages. First, our method has a numerical complexity that grows only linearly with the number of decision makers. Second, our method is independent of the specific decision making test adopted by the SDMs and requires knowledge of only the decision probabilities of the SDMs as a function of time. Third, our method allows for asynchronous decision times among SDMs. To the best of our knowledge, the performance of sequential aggregation schemes for asynchronous decisions has not been previously studied.

Second, we consider the so-called *fastest* and *majority* rules corresponding, respectively, to the decision thresholds $q = 1$ and $q = \lceil N/2 \rceil$. For these rules we provide a comprehensive scalability analysis of both accuracy and decision time. Specifically, in the limit of large group sizes, we provide exact expressions for the expected decision time and the probability of wrong decision for both rules, as a function of the decision probabilities of each SDM. For the *fastest* rule we show that the group decision time converges to the earliest possible decision time of an individual SDM, i.e., the earliest time for which the individual SDM has a non-zero decision probability. Additionally, the *fastest* rule asymptotically obtains the correct answer almost surely, provided the individual SDM is more likely to make the correct decision, rather than the wrong decision, at the earliest possible decision time. For the *majority* rule we show that the probability of wrong decision converges exponentially to zero if the individual SDM has a sufficiently small probability of wrong decision. Additionally, the decision time for the *majority* rule is related to the earliest time at which the individual SDM is more likely to give a decision than to not give a decision. This scalability analysis relies upon novel asymptotic and monotonicity results of certain binomial expansions.

As third main contribution, using our recursive method, we present a comprehensive numerical analysis of sequential decision aggregation based on the *q out of N* rules. As model for the individual SDMs, we adopt the sequential probability ratio test (SPRT), which we characterize as an absorbing Markov chain. First, for the *fastest* and *majority* rules, we report how accuracy and decision time vary as a function of the group size and of the SPRT decision probabilities. Second, in the most general setup, we report how accuracy and decision time vary monotonically as a function of group size and decision threshold. Additionally, we compare the performance of fastest versus majority rules, at fixed group accuracy. We show that the best choice between the fastest rule and the majority rule is a function of group size and group accuracy. Our numerical results illustrate why the design of optimal aggregation rules is a complex task [44]. Finally, we discuss relationships between our analysis of sequential decision aggregation and mental behavior documented in the cognitive psychology and neuroscience literature [45, 11, 12, 13].

Finally, we draw some qualitative lessons about sequential decision aggregation from our mathematical analysis. Surprisingly, our results show that the accuracy of a group is not necessarily improved over the accuracy of an individual. In aggregation based on the *majority* rule, it is true that group accuracy is (exponentially) better than individual accuracy; decision time, however, converges to a constant value for large group sizes. Instead, if a quick decision time is desired, then the *fastest* rule leads, for large group sizes, to decisions being made at the earliest possible time. However, the accuracy of fastest aggregation is not determined by the individual accuracy (i.e., the time integral of the probability of correct decision over time), but is rather determined by the individual accuracy at a specific time instant, i.e., the probability of correct decision at the earliest decision time. Accuracy at this special time might be arbitrarily bad especially for "asymmetric" decision makers (e.g., SPRT with asymmetric thresholds). Arguably, these detailed results for *fastest* and *majority* rules, $q = 1$ and $q = \lfloor N/2 \rfloor$ respectively, are indicative of the accuracy and decision time performance of aggregation rules for small and large thresholds, respectively.

2.1.3 Decision making in cognitive psychology

An additional motivation to study sequential decision aggregation is our interest in sensory information processing systems in the brain. There is a growing belief among neuroscientists [11, 12, 13] that the brain normally engages in an ongoing synthesis of streams of information (stimuli) from multiple sensory modalities. Example modalities include vision, auditory, gustatory, olfactory and somatosensory. While many areas of the brain (e.g., the primary projection pathways) process information from a single sensory modality, many nuclei (e.g., in the Superior Colliculus) are known to receive and integrate stimuli from multiple sensory modalities. Even in these multi-modal sites, a specific stimulus might be dominant. Multi-modal integration is indeed relevant when the response elicited by stimuli from different sensory modalities is statistically different from the response elicited by the most effective of those stimuli presented individually. (Here, the response is quantified in the number of impulses from neurons.) Moreover, regarding data processing in these multi-modal sites, the procedure with which stimuli are processed changes depending upon the intensity of each modality-specific stimulus.

In [11], Werner et al. study a human decision making problem with multiple sensory modalities. They present examples where accuracy and decision time depend upon the strength of the audio and visual components in audio-visual stimuli. They find that, for intact stimuli (i.e., noiseless signals), the decision time improves in multi-modal integration (that is, when both stimuli are simultaneously presented) as compared with uni-sensory integration. Instead, when both stimuli are degraded with noise, multi-modal integration leads to an improvement in both accuracy and decision time. Interestingly, they also identify circumstances for which multi-modal integration leads to performance degradation: performance with an intact stimulus together with a degraded stimulus is sometimes worse than performance with only the intact stimulus.

Another point of debate among cognitive neuroscientists is how to characterize uni-sensory versus multi-modal integration sites. Neuro-physiological studies have traditionally classified as multi-modal sites where stimuli are enhanced, that is, the response to combined stimuli is larger than the sum of the responses to individual stimuli. Recent observations of suppressive responses in multi-modal sites has put this theory to doubt; see [12, 13] and references therein. More specifically, studies have shown that by manipulating the presence and informativeness of stimuli, one can affect the performance (accuracy and decision time) of the subjects in interesting, yet not well understood ways. We envision that a more thorough theoretical understanding of sequential decision aggregation will help bridge the gap between these seemingly contradicting characterization of multi-modal integration sites.

As a final remark about uni-sensory integration sites, it is well known [17] that the cortex in the brain integrates information in *neural groups* by implementing a *drift-diffusion model*. This model is the continuous-time version of the so-called sequential probability ratio test (SPRT) for binary hypothesis testing. We will adopt the SPRT model for our numerical results.

2.1.4 Chapter organization

We start in Section 2.2 by introducing the problem setup. In Section 2.3 we present the numerical method that allows us to analyze the decentralized Sequential Decision Aggregation (SDA) problem; We analyze the two proposed rules in Section 2.4. We also present the numerical results in Section 2.6. Our conclusions are stated in Section 2.7. The appendices contain some results on binomial expansions and on the SPRT.

2.2 Models of sequential aggregation and problem statement

In this section we introduce the model of sequential aggregation and the analysis problem we want to address. Specifically in Subsection 2.2.1 we review the classical sequential binary hypothesis testing problem and the notion of *sequential decision maker*, in Subsection 2.2.2 we define the *q out of N sequential decisions aggregation* setting and, finally, in Subsection 2.2.3, we state the problem we aim to solve.

2.2.1 Sequential decision maker

The classical binary sequential decision problem is posed as follows.

Let H denote a hypothesis which takes on values H_0 and H_1. Assume we are given an individual (called *sequential decision maker (SDM)* hereafter) who repeatedly observes at time $t = 1, 2, \ldots,$ a random variable X taking values in some set $\mathcal{X}$ with the purpose of deciding between H_0 and H_1. Specifically the SDM takes the observations $x(1), x(2), x(3), \ldots$, until it provides its final decision at time τ, which is assumed to be a stopping time for the sigma field sequence generated by the observations, and makes a final decision δ based on the observations up to time τ. The stopping rule together with the final decision rule represent the decision policy of the SDM. The standing assumption is that the conditional joint distributions of the individual observations under each hypothesis are known to the SDM.

In our treatment, we do not specify the type of decision policy adopted by the SDM. A natural way to keep our presentation as general as possible, is to refer to a probabilistic framework that conveniently describes the sequential decision process generated by any decision policy. Specifically, given the decision policy γ, let $\chi_0^{(\gamma)}$ and $\chi_1^{(\gamma)}$ be two random variables defined on the sample space $\mathbb{N} \times \{0,1\} \cup \{?\}$ such that, for $i, j \in \{0,1\}$,

- $\{\chi_j^{(\gamma)} = (t,i)\}$ represents the event that the individual decides in favor of H_i at time t given that the true hypothesis is H_j; and
- $\{\chi_j^{(\gamma)} =?\}$ represents the event that the individual never reaches a decision given that H_j is the correct hypothesis.

Accordingly, define $p_{i|j}^{(\gamma)}(t)$ and $p_{\mathrm{nd}|j}^{(\gamma)}$ to be the probabilities that, respectively, the events $\{\chi_j^{(\gamma)} = (t,i)\}$ and $\{\chi_0^{(\gamma)} =?\}$ occur, i.e,

$$p_{i|j}^{(\gamma)}(t) = \mathbb{P}[\chi_j^{(\gamma)} = (t,i)] \qquad \text{and} \qquad p_{\mathrm{nd}|j}^{(\gamma)} = \mathbb{P}[\chi_j^{(\gamma)} =?].$$

Then the sequential decision process induced by the decision policy γ is completely characterized by the following two sets of probabilities

$$\left\{p_{\mathrm{nd}|0}^{(\gamma)}\right\} \cup \left\{p_{0|0}^{(\gamma)}(t), p_{1|0}^{(\gamma)}(t)\right\}_{t\in\mathbb{N}} \qquad \text{and} \qquad \left\{p_{\mathrm{nd}|1}^{(\gamma)}\right\} \cup \left\{p_{0|1}^{(\gamma)}(t), p_{1|1}^{(\gamma)}(t)\right\}_{t\in\mathbb{N}}, \tag{2.1}$$

where, clearly $p_{\mathrm{nd}|0}^{(\gamma)}+\sum_{t=1}^{\infty}\left(p_{0|0}^{(\gamma)}(t)+p_{1|0}^{(\gamma)}(t)\right)=1$ and $p_{\mathrm{nd}|1}^{(\gamma)}+\sum_{t=1}^{\infty}\left(p_{0|1}^{(\gamma)}(t)+p_{1|1}^{(\gamma)}(t)\right)=1$. In what follows, while referring to a SDM running a sequential distributed hypothesis test with a pre-assigned decision policy, we will assume that the above two probabilities sets are known. From now on, for simplicity, we will drop the superscript (γ).

Together with the probability of no-decision, for $j \in \{0,1\}$ we introduce also the probability of correct decision $p_{\mathrm{c}|j} := \mathbb{P}[\text{say } H_j \,|\, H_j]$ and the probability of wrong decision $p_{\mathrm{w}|j} := \mathbb{P}[\text{say } H_i,\ i \neq j \,|\, H_j]$, that is,

$$p_{\mathrm{c}|j} = \sum_{t=1}^{\infty} p_{j|j}(t) \qquad \text{and} \qquad p_{\mathrm{w}|j} = \sum_{t=1}^{\infty} p_{i|j}(t),\ i \neq j.$$

It is worth remarking that in most of the binary sequential decision making literature, $p_{\mathrm{w}|1}$ and $p_{\mathrm{w}|0}$ are referred as, respectively, the *mis-detection* and *false-alarm* probabilities of error.

Below, we provide a formal definition of two properties that the SDM might or might not satisfy.

Definition 2.2.1 *For a SDM with decision probabilities as in* (2.1)*, the following properties may be defined:*

1. *the SDM has* almost-sure decisions *if, for* $j \in \{0, 1\}$,

$$\sum_{t=1}^{\infty} \left(p_{0|j}(t) + p_{1|j}(t)\right) = 1, \quad \textit{and}$$

2. *the SDM has* finite expected decision time *if, for* $j \in \{0, 1\}$,

$$\sum_{t=1}^{\infty} t \left(p_{0|j}(t) + p_{1|j}(t)\right) < \infty.$$

One can show that the finite expected decision time implies almost-sure decisions.

We conclude this section by briefly discussing examples of sequential decision makers. The classic model is the SPRT model, which we discuss in some detail in the example below and in Section 2.6. Our analysis, however, allows for arbitrary sequential binary hypothesis tests, such as the SPRT with time-varying thresholds [46], constant false alarm rate tests [47], and generalized likelihood ratio tests. Response profiles arise also in neurophysiology, e.g., [48] presents neuron models with a response that varies from unimodal to bimodal depending on the strength of the received stimulus.

Example 2.2.2 (Sequential probability ratio test (SPRT)) *In the case the observations taken are independent, conditioned on each hypothesis, a well-known solution to the above binary decision problem is the so-called* sequential probability ratio test (SPRT) *that we review in Section 2.6. A SDM implementing the SPRT test has both the* almost-sure decisions *and* finite expected decision time *properties.*

Moreover the SPRT test satisfies the following optimality property: among all the sequential tests having pre-assigned values of mis-detection *and* false-alarm *probabilities of error, the SPRT is the test that requires the smallest expected number of iterations for providing a solution.*

In Appendix C we review the methods proposed for computing the probabilities $\{p_{i|j}(t)\}_{t\in\mathbb{N}}$ *when the SPRT test is applied, both in the case* X *is a discrete random variable and in the case* X *is a continuous random variable. For illustration purposes, we provide in Figure 2.1 the probabilities* $p_{i|j}(t)$ *when* $j = 1$ *for the case when* X *is a continuous random variable with a continuous distribution (Gaussian). We also note that* $p_{i|j}(t)$ *might have various interesting distributions.*

2.2.2 The q out of N decentralized hypothesis testing

The basic framework for the binary hypothesis testing problem we analyze in this chapter is the one in which there are N SDMs and one fusion center. The binary hypothesis is denoted by H and it is assumed to take on values H_0 and H_1. Each SDM is assumed to perform individually a binary sequential test; specifically, for $i \in \{1, \dots, N\}$, at time $t \in \mathbb{N}$, SDM i takes the observation $x_i(t)$ on a random variable X_i, defined on some set $\mathcal{X}_i$, and it keeps observing X_i until it provides its decision according to some decision policy γ_i. We assume that

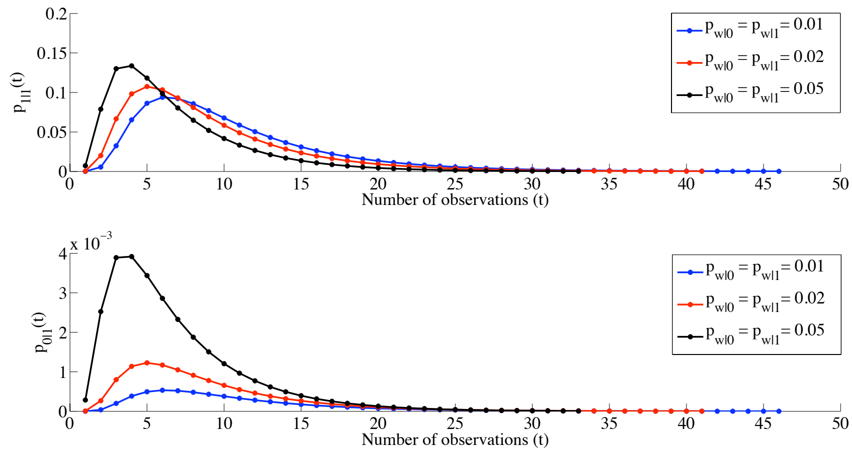


Figure 2.1: This figure illustrates a typical unimodal set of decision probabilities $\{p_{1|1}(t)\}_{t\in\mathbb{N}}$ and $\{p_{0|1}(t)\}_{t\in\mathbb{N}}$. Here the SDM is implementing the sequential probability ratio test with three different accuracy levels (see Section 2.6 for more details).

1. the random variables $\{X_i\}_{i=1}^{N}$ are independent and identically distributed;
2. the SDMs adopt the same decision policy γ, that is, $\gamma_i \cong \gamma$ for all $i \in \{1, \dots, N\}$;
3. the observations taken, conditioned on either hypothesis, are independent from one SDM to another;
4. the conditional joint distributions of the individual observations under each hypothesis are known to the SDMS.

In particular assumptions (i) and (ii) imply that the N decision processes induced by the N SDMs are all described by the same two sets of probabilities

$$\{p_{\mathrm{nd}|0}\} \cup \{p_{0|0}(t), p_{1|0}(t)\}_{t\in\mathbb{N}} \qquad \text{and} \qquad \{p_{\mathrm{nd}|1}\} \cup \{p_{0|1}(t), p_{1|1}(t)\}_{t\in\mathbb{N}}. \quad (2.2)$$

We refer to the above property as *homogeneity* among the SDMs.

Once a SDM arrives to a final local decision, it communicates it to the fusion center. The fusion center collects the messages it receives keeping track of the number of decisions in favor of H_0 and in favor of H_1. A global decision is provided according to a q *out of* N counting rule: roughly speaking, as soon as the hypothesis H_i receives q local decisions in its favor, the fusion center globally decides in favor of H_i. In what follows we refer to the above framework as q *out of* N *sequential decision aggregation* with homogeneous SDMs (denoted as q *out of* N *SDA*, for simplicity).

We describe our setup in more formal terms. Let N denote the size of the group of SDMs and let q be a positive integer such that $1 \leq q \leq N$, then the q *out of* N *SDA* with homogeneous SDMs is defined as follows:

SDMs iteration : For each $i \in \{1, \dots, N\}$, the i-th SDM keeps observing X_i, taking the observations $x_i(1), x_i(2), \dots,$ until time τ_i where it provides its local decision $d_i \in \{0, 1\}$; specifically $d_i = 0$ if it decides in favor of H_0 and $d_i = 1$ if it decides in favor of H_1. The decision d_i is instantaneously communicated (i.e., at time τ_i) to the fusion center.

Fusion center state : The fusion center stores in memory the variables $Count_0$ and $Count_1$, which are initialized to 0, i.e., $Count_0(0) = Count_1(0) = 0$. If at time $t \in \mathbb{N}$ the fusion center has not yet provided a global decision, then it performs two actions in the following order:

(1) it updates the variables $Count_0$ and $Count_1$, according to $Count_0(t) = Count_0(t-1)+n_0(t)$ and $Count_1(t) = Count_1(t-1)+n_1(t)$ where $n_0(t)$ and $n_1(t)$ denote, respectively, the number of decisions equal to 0 and 1 received by the fusion center at time t.

(2) it checks if one of the following two situations is verified

$$(i) \begin{cases} Count_1(t) > Count_0(t), \\ Count_1(t) \geq q, \end{cases} \qquad (ii) \begin{cases} Count_1(t) < Count_0(t). \\ Count_0(t) \geq q. \end{cases} \tag{2.3}$$

If (i) is verified the fusion center globally decides in favor H_1, while if (ii) is verified the fusion center globally decides in favor of H_0. Once the fusion center has provided a global decision the *q out of N SDA* algorithm stops.

Remark 2.2.3 (Notes about SDA) *1. Each SDM has in general a non-zero probability of not giving a decision. In this case, the SDM might keep sampling infinitely without providing any decision to the fusion center.*

2. The fusion center does not need to wait until all the SDM have provided *a decision before a decision is reach on the group level, as one of the two conditions (i) or (ii) in equation 2.3 might be satisfied much before the N SDM provide their decisions.*

3. *While we study in this manuscript the case when a fusion center receives the information from all SDM, we note that a distributed implementation of the SDA algorithm is possible. Analysis similar to the one presented here is possible in that case.* □

2.2.3 Problem formulation

We introduce now some definitions that will be useful throughout this chapter. Given a group of N SDMs running the q *out of* N *SDA* algorithm, $1 \leq q \leq N$, we denote

1. by T the random variable accounting for the number of iterations required to provide a decision

$$T = \min\{t \mid \text{either } \textit{case} \text{ (i) or } \textit{case} \text{ (ii) in equation (2.3) is satisfied}\};$$

2. by $p_{i|j}(t; N, q)$ the probability of deciding, at time t, in favor of H_i given that H_j is correct, i.e.,

$$p_{i|j}(t; N, q) := \mathbb{P}\left[\text{Group of } N \text{ SDMs says } H_i \,|\, H_j, q, T = t\right]; \tag{2.4}$$

3. by $p_{\mathrm{c}|j}(N,q)$ and $p_{\mathrm{w}|j}(N,q)$ the probability of correct decision and of wrong decision, respectively, given that H_j is the correct hypothesis, i.e.,

$$p_{\mathrm{c}|j}(N,q) = \sum_{t=1}^{\infty} p_{j|j}(t;N,q) \qquad \text{and} \qquad p_{\mathrm{w}|j}(N,q) = \sum_{t=1}^{\infty} p_{i|j}(t;N,q), \; i \neq j; \tag{2.5}$$

4. by $p_{\mathrm{nd}|j}(N,q)$, $j \in \{0,1\}$, the probability of no-decision given that H_j is the correct hypothesis, i.e.,

$$p_{\mathrm{nd}|j}(N,q) := 1 - \sum_{t=1}^{\infty} \big(p_{0|j}(t;N,q) + p_{1|j}(t;N,q)\big) = 1 - p_{\mathrm{w}|j}(N,q) - p_{\mathrm{c}|j}(N,q); \tag{2.6}$$

5. by $\mathbb{E}\left[T|H_j,N,q\right]$ the average number of iterations required by the algorithm to provide a decision, given that H_j is the correct hypothesis, i.e.,

$$\mathbb{E}\left[T|H_j,N,q\right] := \begin{cases} \sum_{t=1}^{\infty} t(p_{0|j}(t;N,q) + p_{1|j}(t;N,q)), & \text{if } p_{\mathrm{nd}|j}(N,q) = 0, \\ +\infty, & \text{if } p_{\mathrm{nd}|j}(N,q) > 0. \end{cases} \tag{2.7}$$

Observe that $p_{i|j}(t;1,1)$ coincides with the probability $p_{i|j}(t)$ introduced in (2.1). For ease of notation we will continue using $p_{i|j}(t)$ instead of $p_{i|j}(t;1,1)$.

We are now ready to formulate the problem we aim to solve in this chapter.

Problem 2.2.4 (Sequential decision aggregation) *Consider a group of N homogeneous SDMs with decision probabilities $\{p_{\mathrm{nd}|0}\} \cup \{p_{0|0}(t), p_{1|0}(t)\}_{t\in\mathbb{N}}$ and $\{p_{\mathrm{nd}|1}\} \cup$*

$\left\{p_{0|1}(t), p_{1|1}(t)\right\}_{t\in\mathbb{N}}$. *Assume the* N *SDMs run the* q out of N SDA *algorithm with the purpose of deciding between the hypothesis* H_0 *and* H_1. *For* $j \in \{0,1\}$, *compute the distributions* $\left\{p_{i|j}(t; N, q)\right\}_{t\in\mathbb{N}}$ *as well as the probabilities of correct and wrong decision, i.e.,* $p_{\mathrm{c}|j}(N,q)$ *and* $p_{\mathrm{w}|j}(N,q)$, *the probability of no-decision* $p_{\mathrm{nd}|j}(N,q)$ *and the average number of iterations required to provide a decision, i.e.,* $\mathbb{E}\left[T|H_j, N, q\right]$.

We will focus on the above problem in the next two Sections, both through theoretical and numerical results. Moreover, in Section 2.4, we will concentrate on two particular values of q, specifically for $q = 1$ and $q = \lfloor N/2 \rfloor + 1$, characterizing the tradeoff between the expected decision time, the probabilities of correct and wrong decision and the size of the group of SDMs. When $q = 1$ and $q = \lceil N/2 \rceil$, we will refer to the *q out of N rule* as the *fastest rule* and the *majority rule*, respectively. In this case we will use the following notations

$$p^{(\mathrm{f})}_{\mathrm{c}|j}(N) := p_{\mathrm{c}|j}(N; q = 1), \qquad p^{(\mathrm{f})}_{\mathrm{w}|j}(N) := p_{\mathrm{w}|j}(N; q = 1)$$

and

$$p^{(\mathrm{m})}_{\mathrm{c|j}}(N) := p_{\mathrm{c}|j}(N; q = \lfloor N/2 \rfloor + 1), \qquad p^{(\mathrm{m})}_{\mathrm{w}|j}(N) := p_{\mathrm{w}|j}(N; q = \lfloor N/2 \rfloor + 1).$$

We end this Section by stating two propositions characterizing the *almost-surely decisions* and *finite expected decision time* properties for the group of SDMs.

Proposition 2.2.5 *Consider a group of N SDMs running the q* out of N SDA *algorithm. Let the decision-probabilities of each SDM be as in* (2.2). *For $j \in \{0,1\}$, assume there exists at least one time instant $t_j \in \mathbb{N}$ such that both probabilities $p_{0|j}(t_j)$ and $p_{1|j}(t_j)$ are different from zero. Then the group of SDMs has the* almost-sure decision *property if and only if*

1. *the single SDM has the* almost-sure decision *property;*

2. *N is odd; and*

3. *q is such that $1 \leq q \leq \lceil N/2 \rceil$.*

Proof: First we prove that if the group of SDMs has the *almost-sure decision* property, then properties (i), (ii) and (iii) are satisfied. To do so, we show that if one between the properties (i), (ii) and (iii) fails then there exists an event of probability non-zero that leads the group to not provide a decision. First assume that the single SDM does not have the *almost-sure decision* property, i.e., $p_{\mathrm{nd}|j} > 0$, $j \in \{0,1\}$. Clearly this implies that the event *"all the SDMs of the group do not provide a decision"* has probability of occurring equal to $p_{\mathrm{nd}|j}^N$ which is strictly greater than zero. Second assume that N is even and consider the event *"at time t_j, $N/2$ SDMs decide in favor of H_0 and $N/2$ SDMs decide in favor of H_1"*. Simple combinatorics and probabilistic arguments show that the probability of this event is $\binom{N}{N/2} p_{0|j}^{N/2} p_{1|j}^{N/2}$, which is strictly greater than zero because of the assumption $p_{0|j}(t_j) \neq 0$ and $p_{1|j}(t_j) \neq 0$. Third assume that $q > \lfloor N/2 \rfloor + 1$. In this case we consider the event *"at time t_j, $\lceil N/2 \rceil$ SDMs decide in favor of H_0 and $\lfloor N/2 \rfloor$ SDMs decide in favor of H_1"* that, clearly, leads the group of SDMs to not provide a global decision for any $q > \lfloor N/2 \rfloor + 1$. Similarly to the previous case, we have that the probability of this event is $\binom{N}{\lceil N/2 \rceil} p_{0|j}^{\lceil N/2 \rceil} p_{1|j}^{\lfloor N/2 \rfloor} > 0$.

We prove now that if properties (i), (ii) and (iii) are satisfied then the group of SDMs has the *almost-sure decision* property. Observe that, since each SDM has the *almost-sure decision* property, there exists almost surely a N-tuple $(t_1, \dots, t_N) \in \mathbb{N}^N$ such that the i-th SDM provides its decision at time t_i. Let $\bar{t} := \max\{t_i \mid i \in \{1, \dots, N\}\}$. Since N is odd, then $Count_1(\bar{t}) \neq Count_0(\bar{t})$. Moreover since $q \leq \lfloor N/2 \rfloor + 1$ and $Count_1(\bar{t}) + Count_0(\bar{t}) = N$, either $Count_1(\bar{t}) \geq q$ or $Count_0(\bar{t}) \geq q$ holds true. Hence the fusion center will provide a global decision not later than time $\bar{t}$.

Proposition 2.2.6 *Consider a group of N SDMs running the q* out of *N* SDA *algorithm. Let the decision-probabilities of each SDM be as in* (2.2)*. For $j \in \{0, 1\}$, assume there exists at least one time instant $t_j \in \mathbb{N}$ such that both probabilities $p_{1|j}(t_j)$ and $p_{1|j}(t_j)$ are different from zero. Then the group of SDMs has the* finite expected decision time *property if and only if*

1. *the single SDM has the* finite expected decision time *property;*

2. N is odd; and

3. q is such that $1 \leq q \leq \lceil N/2 \rceil$.

Proof: The proof follows the lines of the proof of the previous proposition.

Remark 2.2.7 *The existence, for $j \in \{0,1\}$, of a time t_j such that $p_{0|j}(t_j) \neq 0$ and $p_{1|j}(t_j) \neq 0$, is necessary only for proving the "if" side of the previous propositions. In other words the validity of properties (i), (ii) and (iii) in Proposition 2.2.5 (resp. in Prop. 2.2.6) guarantees that the group of SDMs possesses the* almost-sure decision *property (resp. the* finite expected decision time *property.)*□

2.3 Recursive analysis of the q-out-of-N sequential aggregation rule

The goal of this section is to provide an efficient method to compute the probabilities $p_{i|j}(t; N, q)$, $i, j \in \{0,1\}$. These probabilities, using equations (2.5), (2.6) and (2.7) will allow us to estimate the probabilities of correct decision, wrong decision and no-decision, as well as the expected number of iterations required to provide the final decision.

We first consider in subsection 2.3.1 the case where $1 \leq q \leq \lfloor N/2 \rfloor$; in subsection 2.3.2 we consider the case where $\lfloor N/2 \rfloor + 1 \leq q \leq N$.

2.3.1 Case $1 \leq q \leq \lfloor N/2 \rfloor$

To present our analysis method, we begin with an informal description of the decision events characterizing the q *out of* N SDA algorithm. Assume that the fusion center provides its decision at time t. This fact implies that neither case (i) nor case (ii) in equation (2.3) has happened at any time before t. Moreover, two distinct set of events may precede time t, depending upon whether the values of the counters $Count_0$ and $Count_1$ at time $t-1$ are smaller than q or not. In a first possible set of events, say the "simple situation," the counters satisfy $0 \leq Count_0(t-1), Count_1(t-1) \leq q-1$ and, hence, the time t is the first time that at least one of the two counters crosses the threshold q. In a second possible set of events, say the "canceling situation," the counters $Count_0(t-1)$ and $Count_1(t-1)$ are greater than q and, therefore, equal. In the canceling situation, there must exist a time instant $\bar{\tau} \leq t-1$ such that $Count_0(\bar{\tau}-1) < q$, $Count_1(\bar{\tau}-1) < q$ and $Count_0(\tau) = Count_1(\tau) \geq q$ for all $\tau \in \{\bar{\tau}+1, \dots, t-1\}$. In other words, both counters cross the threshold q at the same time instant $\bar{\tau}$ reaching the same value, that is, $Count_0(\bar{\tau}) = Count_1(\bar{\tau})$, and, for time $\tau \in \{\bar{\tau}+1, \dots, t-1\}$, the number $n_0(\tau)$ of SDMs deciding in favor of H_0 at time τ and the number $n_1(\tau)$ of SDMs deciding in favor of H_1 at time τ cancel each other out, that is, $n_0(\tau) = n_1(\tau)$.

In what follows we study the probability of the simple and canceling situations. To keep track of both possible set of events, we introduce four probability functions, α, β, $\bar{\alpha}$, $\bar{\beta}$. The functions α and β characterize the simple situation, while $\bar{\alpha}$ and $\bar{\beta}$ characterize the canceling situation. First, for the simple situation, define the probability function $\alpha : \mathbb{N} \times \{0, \dots, q-1\} \times \{0, \dots, q-1\} \to [0, 1]$ as follows: given a group of $s_0 + s_1$ SDMs, $\alpha(t, s_0, s_1)$ is the probability that

1. all the $s_0 + s_1$ SDMs have provided a decision up to time t; and

2. considering the variables $Count_0$ and $Count_1$ restricted to this group of $s_0 + s_1$ SDMs , $Count_0(t) = s_0$ and $Count_1(t) = s_1$.

Also, define the probability function $\beta_{1|j} : \mathbb{N} \times \{0, \dots, q-1\} \times \{0, \dots, q-1\} \to [0, 1]$, $j \in \{0, 1\}$ as follows: given a group of $N - (s_0 + s_1)$ SDMs, $\beta_{1|j}(t, s_0, s_1)$ is the probability that

1. no SDMs have provided a decision up to time $t - 1$; and

2. considering the variables $Count_0$ and $Count_1$ restricted to this group of $N-(s_0+s_1)$ SDMs, $Count_0(t)+s_0 < Count_1(t)+s_1$, and $Count_1(t)+s_1 \geq q$.

Similarly, it is straightforward to define the probabilities $\beta_{0|j}$, $j \in \{0, 1\}$.

Second, for the canceling situation, define the probability function $\bar{\alpha} : \mathbb{N} \times \{q, \dots, \lfloor N/2 \rfloor\} \to [0, 1]$ as follows: given a group of $2s$ SDMs, $\bar{\alpha}(t, s)$ is the probability that

1. all the $2s$ SDMs have provided a decision up to time t; and

2. there exists $\bar{\tau} \leq t$ such that, considering the variables $Count_0$ and $Count_1$ restricted to this group of $2s$ SDMs

   - $Count_0(\bar{\tau}-1) < q$ and $Count_1(\bar{\tau}-1) < q$;

   - $Count_0(\tau) = Count_1(\tau) \geq q$ for all $\tau \geq \bar{\tau}$.

Also, define the probability function $\bar{\beta}_{1|j} : \mathbb{N} \times \{q, \dots \lfloor N/2 \rfloor\} \rightarrow [0,1]$, $j \in \{0,1\}$ as follows: given a group of $N-2s$ SDMs, $\bar{\beta}_{1|j}(t,s)$ is the probability that

1. no SDMs have provided a decision up to time $t-1$; and

2. at time t the number of SDMs providing a decision in favor of H_1 is strictly greater of the number of SDMs providing a decision in favor of H_0.

Similarly, it is straightforward to define the probabilities $\bar{\beta}_{0|j}$, $j \in \{0,1\}$.

Note that, for simplicity, we do not explicitly keep track of the dependence of the probabilities β and $\bar{\beta}$ upon the numbers N and q. The following proposition shows how to compute the probabilities $\left\{p_{i|j}(t;N,q)\right\}_{t=1}^{\infty}$, $i,j \in \{0,1\}$, starting from the above definitions.

Proposition 2.3.1 ***(q out of N: a recursive formula)*** *Consider a group of N SDMs, running the* q out of N SDA *algorithm. Without loss of generality,* *assume H_1 is the correct hypothesis. Then, for $i \in \{0,1\}$, we have, for $t = 1$,*

$$p_{i|1}(1; N, q) = \beta_{i|1}(1, 0, 0), \tag{2.8}$$

and, for $t \geq 2$,

$$p_{i|1}(t; N, q) = \sum_{s_0=0}^{q-1} \sum_{s_1=0}^{q-1} \binom{N}{s_1 + s_0} \alpha(t-1, s_0, s_1) \beta_{i|1}(t, s_0, s_1) \tag{2.9}$$

$$+ \sum_{s=q}^{\lfloor N/2 \rfloor} \binom{N}{2s} \bar{\alpha}(t-1, s) \bar{\beta}_{i|1}(t, s). \tag{2.10}$$

Proof: The proof that the formula in (2.8) hold true follows trivially form the definition of the quantities $\beta_{1|1}(1,0,0)$ and $\beta_{0|1}(1,0,0)$. We start by providing three useful definitions.

First, let E_t denote the event that the SDA with the q *out of* N rule provides its decision at time t in favor of H_1.

Second, for s_0 and s_1 such that $0 \leq s_0, s_1 \leq q-1$, let $E_{s_0,s_1,t}$ denote the event such that

1. there are s_0 SDMs that have decided in favor of H_0 up to time $t-1$;

2. there are s_1 SDMs that have decided in favor of H_1 up to time $t-1$;

3. there exist two positive integer number r_0 and r_1 such that

   - $s_0 + r_0 < s_1 + r_1$ and $s_1 + r_1 \geq q$.

- at time t, r_0 SDMs decides in favor of H_0 while r_1 SDMs decides in favor of H_1

Third, for $q \leq s \leq \lfloor N/2 \rfloor$, let $E_{s,t}$ denote the event such that

1. $2s$ SDMs have provided their decision up to time $t-1$ balancing their decision, i.e., there exists $\bar{\tau} \leq t-1$ with the properties that, considering the variables $Count_-$ and $Count_+$ restricted to these $2s$ SDMs
   - $Count_0(\tau) < q$, $Count_1(\tau) < q$, for $1 \leq \tau \leq \bar{\tau} - 1$;
   - $Count_0(\tau) = Count_1(\tau)$ for $\bar{\tau} \leq \tau \leq t-1$;
   - $Count_0(t-1) = Count_1(t-1) = s$.
2. at time t the number of SDMs providing their decision in favor of H_1 is strictly greater than the number of SDMs deciding in favor of H_0.

Observe that

$$E_t = \left(\bigcup_{0 \leq s_0, s_1 \leq q-1} E_{s_0,s_1,t} \right) \bigcup \left(\bigcup_{q \leq s \leq \lfloor N/2 \rfloor} E_{s,t} \right).$$

Since $E_{s_0,s_1,t}$, $0 \leq s_0, s_1 \leq q-1$, and $E_{s,t}$, $q \leq s \leq \lfloor N/2 \rfloor$ are disjoint sets, we can write

$$\mathbb{P}[E_t] = \sum_{0 \leq s_0, s_1 \leq q-1} \mathbb{P}[E_{s_0,s_1,t}] + \sum_{q \leq s \leq \lfloor N/2 \rfloor} \mathbb{P}[E_{s,t}]. \tag{2.11}$$

Observe that, according to the definitions of $\alpha(t-1, s_0, s_1)$, $\bar{\alpha}(t-1, s)$, $\beta_{1|1}(t, s_0, s_1)$ and $\bar{\beta}_{1|1}(t, s)$, provided above,

$$\mathbb{P}\left[E_{s_0,s_1,t}\right] = \binom{N}{s_1 + s_0} \alpha(t-1, s_0, s_1)\beta_{1|1}(t, s_0, s_1) \tag{2.12}$$

and that

$$\mathbb{P}\left[E_{s,t}\right] = \binom{N}{2s} \bar{\alpha}(t-1, s)\bar{\beta}_{1|1}(t, s). \tag{2.13}$$

Plugging equations (2.12) and (2.13) into equation (2.11) concludes the proof of the Theorem. Formulas, similar to the ones in (2.8) and (2.9) can be provided for computing also the probabilities $\left\{p_{i|0}(t; N, q)\right\}_{t=1}^{\infty}$, $i \in \{0, 1\}$.

As far as the probabilities $\alpha(t, s_0, s_1)$, $\bar{\alpha}(t, s)$, $\beta_{i|j}(t, s_0, s_1)$, $\bar{\beta}_{i|j}(t, s)$, $i, j \in \{0, 1\}$, are concerned, we now provide expressions to calculate them.

Proposition 2.3.2 *Consider a group of N SDMs, running the q* out of *N* SDA *algorithm for $1 \leq q \leq \lfloor N/2 \rfloor$. Without loss of generality, assume H_1 is the correct hypothesis. For $i \in \{0, 1\}$, let $\pi_{i|1} : \mathbb{N} \to [0, 1]$ denote the cumulative probability up to time t that a single SDM provides the decision H_i, given that H_1 is the correct hypothesis, i.e.,*

$$\pi_{i|1}(t) = \sum_{s=1}^{t} p_{i|1}(t). \tag{2.14}$$

For $t \in \mathbb{N}$, $s_0, s_1 \in \{1, \ldots, q-1\}$, $s \in \{q, \ldots, \lfloor N/2 \rfloor\}$, the probabilities $\alpha(t, s_0, s_1)$, $\bar{\alpha}(t, s)$, $\beta_{1|1}(t, s_0, s_1)$, and $\bar{\beta}_{1|1}(t, s)$ satisfy the following relationships (explicit for

α, β, $\bar{\beta}$ and recursive for $\bar{\alpha}$):

$$\alpha(t, s_0, s_1) = \binom{s_0 + s_1}{s_0} \pi_{0|1}^{s_0}(t) \pi_{1|1}^{s_1}(t),$$

$$\begin{aligned}\bar{\alpha}(t, s) = &\sum_{s_0=0}^{q-1} \sum_{s_1=0}^{q-1} \binom{2s}{s_0 + s_1} \binom{2s - s_0 - s_1}{s - s_0} \alpha(t-1, s_0, s_1) p_{0|1}^{s-s_0}(t) p_{1|1}^{s-s_1}(t) \\ &+ \sum_{h=q}^{s} \binom{2s}{2h} \binom{2s - 2h}{s - h} \bar{\alpha}(t-1, h) p_{0|1}^{s-h}(t) p_{1|1}^{s-h}(t),\end{aligned}$$

$$\begin{aligned}\beta_{1|1}(t, s_0, s_1) = &\sum_{h_1=q-s_1}^{N-\bar{s}} \binom{N - \bar{s}}{h_1} p_{1|1}^{h_1}(t) \\ &\times \left[\sum_{h_0=0}^{m} \binom{N - \bar{s} - h_1}{h_0} p_{0|1}^{h_0}(t) \left(1 - \pi_{1|1}(t) - \pi_{0|1}(t)\right)^{N-\bar{s}-h_0-h_1} \right],\end{aligned}$$

$$\begin{aligned}\bar{\beta}_{1|1}(t, s) = &\sum_{h_1=1}^{N-2s} \binom{N - 2s}{h_1} p_{1|1}^{h_1}(t) \\ &\times \left[\sum_{h_0=0}^{\bar{m}} \binom{N - 2s - h_1}{h_0} p_{0|1}^{h_0}(t) (1 - \pi_{1|1}(t) - \pi_{0|1}(t))^{N-2s-h_0-h_1} \right],\end{aligned}$$

where $\bar{s} = s_0 + s_1$, $m = \min\{h_1 + s_1 - s_0 - 1, N - (s_0 + s_1) - h_1\}$ and $\bar{m} = \min\{h_1 - 1, N - 2s - h_1\}$. Moreover, corresponding relationships for $\beta_{0|1}(t, s_0, s_1)$ and $\bar{\beta}_{0|1}(t, s)$ are obtained by exchanging the roles of $p_{1|1}(t)$ with $p_{0|1}(t)$ in the relationships for $\beta_{1|1}(t, s_0, s_1)$ and $\bar{\beta}_{1|1}(t, s)$.

Proof: The evaluation of $\alpha(t, s_0, s_1)$ follows from standard probabilistic arguments. Indeed, observe that, given a first group of s_0 SDMs and a second group of s_1 SDMs, the probability that all the SDMs of the first group have decided in favor of H_0 up to time t and all the SDMs of the second group have decided in favor of H_1 up to time t is given by $\pi_{0|1}^{s_0}(t)\pi_{1|1}^{s_1}(t)$. The desired result follows from the fact that there are $\binom{s_1+s_0}{s_0}$ ways of dividing a group of s_0+s_1 SDMs into two subgroups of s_0 and s_1 SDMs.

Consider now $\bar{\alpha}(t,s)$. Let $E_{\bar{\alpha}(t,s)}$ denote the event of which $\bar{\alpha}(t,s)$ is the probability of occurring, that is, the event that, given a group of $2s$ SDMs,

1. all the $2s$ SDMs have provided a decision up to time t; and
2. there exists $\bar{\tau} \leq t$ such that, considering the variables $Count_0$ and $Count_1$ restricted to this group of $2s$ SDMs
   - $Count_0(\bar{\tau}-1) < q$ and $Count_1(\bar{\tau}-1) < q$;
   - $Count_0(\tau) = Count_1(\tau) \geq q$ for all $\tau \geq \bar{\tau}$.

Now, for a group of $2s$ SDMs, for $0 \leq s_0, s_1 \leq q-1$, let E_{t-1,s_0,s_1} denote the event that

1. s_0 (resp. s_1) SDMs have decided in favor of H_0 (resp. H_1) up to time $t-1$;
2. $s-s_0$ (resp. $s-s_1$) SDMs decide in favor of H_0 (resp. H_1) at time t.

Observing that for s_0+s_1 assigned SDMs the probability that fact (i) is verified is given by $\alpha(t-1, s_0, s_1)$ we can write that

$$\mathbb{P}[E_{t-1,s_0,s_1}] = \binom{2s}{s_0+s_1}\binom{2s-s_0-s_1}{s-s_0}\alpha(t-1,s_0,s_1)p_{0|1}^{s-s_0}(t)p_{1|1}^{s-s_1}(t).$$

Consider again a group of $2s$ SDMs and for $q \leq h \leq s$ let $\bar{E}_{t-1,h}$ denote the event that

1. $2h$ SDMs have provided a decision up to time $t-1$;
2. there exists $\bar{\tau} \leq t-1$ such that, considering the variables $Count_0$ and $Count_1$ restricted to the group of $2h$ SDMs that have already provided a decision,
   - $Count_0(\bar{\tau}-1) < q$ and $Count_1(\bar{\tau}-1) < q$;
   - $Count_0(\tau) = Count_1(\tau) \geq q$ for all $\tau \geq \bar{\tau}$; and
   - $Count_0(t-1) = Count_1(t-1) = h$;
3. at time instant t, $s-h$ SDMs decide in favor of H_0 and $s-h$ SDMs decide in favor of H_1.

Observing that for $2h$ assigned SDMs the probability that fact (i) and fact (ii) are verified is given by $\bar{\alpha}(t-1,h)$, we can write that

$$\mathbb{P}[\bar{E}_{t-1,h}] = \binom{2s}{2h}\binom{2s-2h}{s-h}\bar{\alpha}(t-1,h)p_{0|1}^{s-h}(t)p_{1|1}^{s-h}(t).$$

Observe that

$$E_{\bar{\alpha}(t,s)} = \left(\bigcup_{s_0=0}^{q}\bigcup_{s_1=0}^{q} E_{t-1,s_0,s_1}\right)\bigcup\left(\bigcup_{h=q}^{\lfloor N/2\rfloor}\bar{E}_{t-1,h}\right).$$

Since the events E_{t-1,s_0,s_1}, $0 \leq s_0, s_1 < q$ and $\bar{E}_{t-1,h}$, $q \leq h \leq \lfloor N/2 \rfloor$, are all disjoint we have that

$$\mathbb{P}[E_{\bar{\alpha}(t,s)}] = \sum_{s_0=0}^{q-1} \sum_{s_1=0}^{q-1} \mathbb{P}[E_{t-1,s_0,s_1}] + \sum_{h=q}^{s} \mathbb{P}[\bar{E}_{t-1,h}].$$

Plugging the expressions of $\mathbb{P}[E_{t-1,s_0,s_1}]$ and $\mathbb{P}[\bar{E}_{t-1,h}]$ in the above equality gives the recursive relationship for computing $\bar{\alpha}(t,s)$.

Consider now the probability $\beta_{1|1}(t, s_0, s_1)$. Recall that this probability refers to a group of $N-(s_0+s_1)$ SDMs. Let us introduce some notations. Let $E_{\beta_{1|1}(t,s_0,s_1)}$ denote the event of which $\beta_{1|1}(t, s_0, s_1)$ represents the probability of occurring and let $E_{t;h_1,s_1,h_0,s_0}$ denote the event that, at time t

- h_1 SDMs decides in favor of H_1;
- h_0 SDMs decides in favor of H_0;
- the remaining $N-(s_0+s_1)-(h_0+h_1)$ do not provide a decision up to time t.

Observe that the above event is well-defined if and only if $h_0+h_1 \leq N-(s_0+s_1)$. Moreover $E_{t;h_1,s_1,h_0,s_0}$ contributes to $\beta_{1|1}(t, s_0, s_1)$, i.e., $E_{t;h_1,s_1,h_0,s_0} \subseteq E_{\beta_{1|1}(t,s_0,s_1)}$ if and only if $h_1 \geq q - s_1$ and $h_0 < h_1 + s_1 - s_0$ (the necessity of these two inequalities follows directly from the definition of $\beta_{1|1}(t, s_0, s_1)$). Considering the three inequalities $h_0 + h_1 \leq N - (s_0 + s_1)$, $h_1 \geq q - s_1$ and $h_0 < h_1 + s_1 - s_0$, it follows that

$$E_{\beta_{1|1}(t,s_0,s_1)} = \bigcup \left\{ E_{t;h_1,s_1,h_0,s_0} \mid q - s_1 \leq h_1 \leq N - (s_0 + s_1) \ \text{ and } \ h_0 \leq m \right\},$$

where $m = \min\{h_1 + s_1 - s_0 - 1, N - (s_0 + s_1) - h_1\}$. To conclude it suffices to observe that the events $E_{t;h_1,s_1,h_0,s_0}$ for $q - s_1 \leq h_1 \leq N - (s_0 + s_1)$ and $h_0 \leq m$ are disjoint events and that

$$\mathbb{P}[E_{t;h_1,s_1,h_0,s_0}] = \binom{N - \bar{s}}{j} p_{1|1}^{h_1}(t) \binom{N - \bar{s} - h_1}{h_0} p_{0|1}^{h_0}(t) \left(1 - \pi_{1|1}(t) - \pi_{0|1}(t)\right)^{N - \bar{s} - h_0 - h_1},$$

where $\bar{s} = s_0 + s_1$.

The probability $\bar{\beta}_{1|1}(t,s)$ can be computed reasoning similarly to $\beta_{1|1}(t,s_0,s_1)$.

Now we describe some properties of the above expressions in order to assess the computational complexity required by the formulas introduced in Proposition 2.3.1 in order to compute $\left\{p_{i|j}(t;N,q)\right\}_{t=1}^{\infty}$, $i,j \in \{0,1\}$. From the expressions in Proposition 2.3.2 we observe that

- $\alpha(t,s_0,s_1)$ is a function of $\pi_{0|1}(t)$ and $\pi_{1|1}(t)$;
- $\bar{\alpha}(t,s)$ is a function of $\alpha(t-1,s_0,s_1)$, $0 \leq s_0, s_1 \leq q-1$, $p_{0|1}(t)$, $p_{1|1}(t)$ and $\bar{\alpha}(t-1,h)$, $q \leq h \leq s$;
- $\beta_{i|1}(t,s_0,s_1)$, $\bar{\beta}_{i|1}$, $i \in \{0,1\}$, are functions of $p_{0|1}(t)$, $p_{1|1}(t)$, $\pi_{0|1}(t)$ and $\pi_{1|1}(t)$.

Moreover from equation (2.14) we have that $\pi_{i|j}(t)$ is a function of $\pi_{i|j}(t-1)$ and $p_{i|j}(t)$.

Based on the above observations, we deduce that $p_{0|1}(t; N, q)$ and $p_{1|1}(t; N, q)$ can be seen as the output of a dynamical system having the $(\lfloor N/2 \rfloor - q + 3)$-th dimensional vector with components the variables $\pi_{0|1}(t-1)$, $\pi_{1|1}(t-1)$, $\bar{\alpha}(t-1, s)$, $q \leq h \leq \lfloor N/2 \rfloor$ as states and the two dimensional vector with components $p_{0|1}(t)$, $p_{1|1}(t)$, as inputs. As a consequence, it follows that the iterative method we propose to compute $\left\{p_{i|j}(t; N, q)\right\}_{t=1}^{\infty}$, $i, j \in \{0, 1\}$, requires keeping in memory a number of variables which grows linearly with the number of SDMs.

2.3.2 Case $\lfloor N/2 \rfloor + 1 \leq q \leq N$

The probabilities $p_{i|j}(t; N, q)$, $i, j \in \{0, 1\}$ in the case where $\lfloor N/2 \rfloor + 1 \leq q \leq N$ can be computed according to the expressions reported in the following Proposition.

Proposition 2.3.3 *Consider a group of N SDMs, running the q* out of N SDA *algorithm for $\lfloor N/2 \rfloor + 1 \leq q \leq N$. Without loss of generality, assume H_1 is the correct hypothesis. For $i \in \{0, 1\}$, let $\pi_{i|1} : \mathbb{N} \to [0, 1]$ be defined as* (2.14)*. Then, for $i \in \{0, 1\}$, we have for $t = 1$*

$$p_{i|1}(1; N, q) = \sum_{h=q}^{N} \binom{N}{h} p_{i|1}^{h}(1) \left(1 - p_{i|1}(1)\right)^{N-h} \tag{2.15}$$

and for $t \geq 2$

$$p_{i|1}(t;N,q) = \sum_{k=0}^{q-1} \binom{N}{k} \pi_{i|1}^{k}(t-1) \sum_{h=q-k}^{N-k} \binom{N-k}{h} p_{i|1}^{h}(t) \left(1-\pi_{i|1}(t)\right)^{N-(h+k)}. \tag{2.16}$$

Proof: Let $t = 1$. Since $q > N/2$, the probability that the fusion center decides in favor of H_i at time $t = 1$ is given by the probability that al least q SDMs decide in favor of H_i at time 1. From standard combinatorics arguments this probability is given by (2.15).

If $t > 1$, the probability that the fusion center decides in favor of H_i at time t is given by the probability that h SDMs, $0 \leq h < q$, have decided in favor of H_i up to time $t-1$, and that at least $q-h$ SDMs decide in favor of H_i at time t. Formally let $E_t^{(i)}$ denote the event that the fusion center provides its decision in favor of H_i at time t and let $E_{h,t;k,t-1}^{(i)}$ denote the event that k SDMs have decided in favor of H_i up to time $t-1$ and h SDMs decide in favor of H_i at time t. Observe that

$$E_t^{(i)} = \bigcup_{k=0}^{q-1} \bigcup_{h=q-k}^{N-k} E_{h,t;k,t-1}^{(i)}.$$

Since $E_{h,t;k,t-1}^{(i)}$ are disjoint sets it follows that

$$\mathbb{P}\left[E_t^{(i)}\right] = \sum_{k=0}^{q-1} \sum_{h=q-k}^{N-k} \mathbb{P}\left[E_{h,t;k,t-1}^{(i)}\right].$$

The proof is concluded by observing that

$$\mathbb{P}\left[E^{(i)}_{h,t;k,t-1}\right] = \binom{N}{k}\pi^k_{i|1}(t-1)\binom{N-k}{h}p^h_{i|1}(t)\left(1-\pi_{i|1}(t)\right)^{N-(h+k)}.$$

Regarding the complexity of the expressions in (2.16) it is easy to see that the probabilities $p_{i|j}(t; N, q)$, $i, j \in \{0, 1\}$ can be computed as the output of a dynamical system having the two dimensional vector with components $\pi_{0|1}(t-1), \pi_{1|1}(t-1)$ as state and the two dimensional vector with components $p_{0|1}(t), p_{1|1}(t)$ as input. In this case the dimension of the system describing the evolution of the desired probabilities is independent of N.

2.4 Scalability analysis of the fastest and majority sequential aggregation rules

The goal of this section is to provide some theoretical results characterizing the probabilities of being correct and wrong for a group implementing the *q-out-of-N* SDA rule. We also aim to characterize the probability with which such a group fails to reach a decision in addition to the time it takes for this group to stop running any test. In Sections 2.4.1 and 2.5.1 we consider the fastest and the majority rules, namely the thresholds $q = 1$ and $q = \lceil N/2 \rceil$, respectively;

we analyze how these two counting rules behave for increasing values of N. In Section 2.5.2, we study how these quantities vary with arbitrary values q and fixed values of N.

2.4.1 The fastest rule for varying values of N

In this section we provide interesting characterizations of accuracy and expected time under the *fastest* rule, i.e., the counting rules with threshold $q = 1$. For simplicity we restrict to the case where the group has the *almost-sure* decision property. In particular we assume the following two properties.

Assumption 2.5 *The number N of SDMs is odd and the SDMs satisfy the* almost-sure *decision property.*

Here is the main result of this subsection. Recall that $p^{(\mathrm{f})}_{\mathrm{w}|1}(N)$ is the probability of wrong decision by a group of N SDMs implementing the fastest rule (assuming H_1 is the correct hypothesis).

Proposition 2.5.1 (Accuracy and expected time under the fastest rule) *Consider the q* out of *N* SDA *algorithm under Assumption 2.5. Assume $q = 1$, that is, adopt the* fastest *SDA rule. Without loss of generality, assume H_1 is the correct hypothesis. Define the* earliest possible decision time

$$\bar{t} := \min\{t \in \mathbb{N} \mid \textit{either } p_{1|1}(t) \neq 0 \textit{ or } p_{0|1}(t) \neq 0\}. \tag{2.17}$$

Then the probability of error satisfies

$$\lim_{N\to\infty} p^{(\mathrm{f})}_{\mathrm{w}|1}(N) = \begin{cases} 0, & \textit{if } p_{1|1}(\bar{t}) > p_{0|1}(\bar{t}), \\ 1, & \textit{if } p_{1|1}(\bar{t}) < p_{0|1}(\bar{t}), \\ \frac{1}{2}, & \textit{if } p_{1|1}(\bar{t}) = p_{0|1}(\bar{t}), \end{cases} \tag{2.18}$$

and the expected decision time satisfies

$$\lim_{N\to\infty} \mathbb{E}\left[T|H_1, N, q=1\right] = \bar{t}. \tag{2.19}$$

Proof: We start by observing that in the case where the fastest rule is applied, formulas in (2.9) simplifies to

$$p_{1|1}(t; N, q=1) = \beta_{1|1}(t, 0, 0), \qquad \text{for all } t \in \mathbb{N}.$$

Now, since $p_{1|1}(t) = p_{0|1}(t) = 0$ for $t < \bar{t}$, it follows that

$$p_{1|1}(t; N, q=1) = \beta_{1|1}(t, 0, 0) = 0, \qquad t < \bar{t}.$$

Moreover we have $\pi_{1|1}(\bar{t}) = p_{1|1}(\bar{t})$ and $\pi_{0|1}(\bar{t}) = p_{0|1}(\bar{t})$. According to the definition of the probability $\beta_{1|1}(\bar{t}, 0, 0)$, we write

$$\beta_{1|1}(\bar{t}, 0, 0) = \sum_{j=1}^{N} \binom{N}{j} p^{j}_{1|1}(\bar{t}) \left\{ \sum_{i=0}^{m} \binom{N-j}{i} p^{i}_{0|1}(\bar{t}) \left(1 - p_{1|1}(\bar{t}) - p_{0|1}(\bar{t})\right)^{N-i-j} \right\},$$

where $m = \min\{j-1, N-j\}$, or equivalently

$$\begin{aligned}
\beta_{1|1}(\bar{t},0,0) = & \sum_{j=1}^{\lfloor N/2 \rfloor} \binom{N}{j} p_{1|1}^{j}(\bar{t}) \left\{ \sum_{i=0}^{j-1} \binom{N-j}{i} p_{0|1}^{i}(\bar{t}) \left(1 - p_{1|1}(\bar{t}) - p_{0|1}(\bar{t})\right)^{N-i-j} \right\} \\
& + \sum_{j=\lceil N/2 \rceil}^{N} \binom{N}{j} p_{1|1}^{j}(\bar{t}) \left\{ \sum_{i=0}^{N-j} \binom{N-j}{i} p_{0|1}^{i}(\bar{t}) \left(1 - p_{1|1}(\bar{t}) - p_{0|1}(\bar{t})\right)^{N-i-j} \right\} \\
= & \sum_{j=1}^{\lfloor N/2 \rfloor} \binom{N}{j} p_{1|1}^{j}(\bar{t}) \left\{ \sum_{i=0}^{j-1} \binom{N-j}{i} p_{0|1}^{i}(\bar{t}) \left(1 - p_{1|1}(\bar{t}) - p_{0|1}(\bar{t})\right)^{N-i-j} \right\} \\
& + \sum_{j=\lceil N/2 \rceil}^{N} \binom{N}{j} p_{1|1}^{j}(\bar{t}) \left(1 - p_{1|1}(\bar{t})\right)^{N-j}. \qquad (2.20)
\end{aligned}$$

An analogous expression for $\beta_{0|1}(\bar{t},0,0)$ can be obtained by exchanging the roles of $p_{0|1}(\bar{t})$ and $p_{0|1}(\bar{t})$ in equation (2.20). The rest of the proof is articulated as follows. First, we prove that

$$\lim_{N\to\infty} \left(p_{1|1}(\bar{t}; N, q=1) + p_{0|1}(\bar{t}; N, q=1)\right) = \lim_{N\to\infty} \left(\beta_{1|1}(\bar{t},0,0) + \beta_{0|1}(\bar{t},0,0)\right) = 1. \qquad (2.21)$$

This fact implies that equation (2.19) holds and that, if $p_{1|1}(\bar{t}) = p_{0|1}(\bar{t})$, then $\lim_{N\to\infty} p_{\mathrm{w}|1}^{(\mathrm{f})}(N) = 1/2$. Indeed

$$\lim_{N\to\infty} \mathbb{E}\left[T | H_j, N, q=1\right] = \lim_{N\to\infty} \sum_{t=1}^{\infty} t(p_{0|j}(t; N, q=1) + p_{i|j}(t; N, q=1)) = \bar{t}.$$

Moreover, if $p_{1|1}(\bar{t}) = p_{0|1}(\bar{t})$, then also $(\beta_{1|1}(\bar{t},0,0) = \beta_{0|1}(\bar{t},0,0)$.

Second, we prove that $p_{1|1}(\bar{t}) > p_{0|1}(\bar{t})$ implies $\lim_{N\to\infty} \beta_{0|1}(\bar{t},0,0) = 0$. As a consequence, we have that $\lim_{N\to\infty} \beta_{1|1}(\bar{t},0,0) = 1$ or equivalently that $\lim_{N\to\infty} p_{\mathrm{w}|1}^{(\mathrm{f})}(N) = 0$.

To show equation (2.21), we consider the event *the group is not giving the decision at time* $\bar{t}$. We aim to show that the probability of this event goes to zero as $N \to \infty$. Indeed we have that

$$\mathbb{P}\left[T \neq \bar{t}\right] = \mathbb{P}\left[T > \bar{t}\right] = 1 - \left(p_{1|1}(\bar{t}, N) + p_{0|1}(\bar{t}, N)\right),$$

and, hence, $\mathbb{P}\left[T > \bar{t}\right] = 0$ implies $p_{1|1}(\bar{t}, N) + p_{0|1}(\bar{t}, N) = 1$. Observe that

$$\mathbb{P}\left[T > \bar{t}\right] = \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{2j} \binom{2j}{j} p_{i|1}(\bar{t})^j p_{0|i}(\bar{t})^j \Big(1 - p_{i|1}(\bar{t}) - p_{0|i}(\bar{t})\Big)^{N-2j}.$$

For simplicity of notation, let us denote $x := p_{0|1}(\bar{t})$ and $y := p_{0|1}(\bar{t})$. We distinguish two cases, (i) $x \neq y$ and (ii) $x = y$.

Case $x \neq y$. We show that in this case there exists $\bar{\epsilon} > 0$, depending only on x and y, such that

$$\binom{2j}{j} x^j y^j < (x + y - \bar{\epsilon})^{2j}, \quad \text{for all } j \geq 1. \tag{2.22}$$

First of all observe that, since $\binom{2j}{j} x^j y^j$ is just one term of the Newton binomial expansion of $(x+y)^{2j}$, we know that $\binom{2j}{j} x^j y^j < (x+y)^{2j}$ for all $j \in \mathbb{N}$. Define $\epsilon(j) := x + y - \binom{2j}{j}^{1/2j} \sqrt{xy}$ and observe that proving equation (2.22) is equivalent to proving $\lim_{j\to\infty} \epsilon(j) > 0$. Indeed if $\lim_{j\to\infty} \epsilon(j) > 0$, then $\inf_{j\in\mathbb{N}} \epsilon(j) > 0$ and thereby we can define $\bar{\epsilon} := \inf_{j\in\mathbb{N}} \epsilon(j)$. To prove the inequality $\lim_{j\to\infty} \epsilon(j) > 0$, let us compute $\lim_{j\to\infty} \binom{2j}{j}^{1/(2j)}$. By applying Stirling's formula we can write

$$\lim_{j\to\infty} \binom{2j}{j}^{1/(2j)} = \lim_{j\to\infty} \left(\frac{\sqrt{2\pi 2j}\, \left(\frac{2j}{e}\right)^{2j}}{2\pi j \left(\frac{j}{e}\right)^{2j}} \right)^{1/(2j)} = \left(\sqrt{\frac{1}{\pi j^2}} 2^{2j} \right)^{1/(2j)} = 2$$

and, in turn, $\lim_{j\to\infty} \epsilon(j) = x+y-2\sqrt{xy}$. Clearly, if $x \neq y$, then $x+y-2\sqrt{xy} > 0$. Defining $\bar{\epsilon} := \inf_{j\in\mathbb{N}} \epsilon(j)$, we can write

$$\begin{aligned}
&\lim_{N\to\infty} \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{2j}\binom{2j}{j} x^j y^j \left(1-x-y\right)^{N-2j} \\
&\leq \lim_{N\to\infty} \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{2j} \left(x+y-\bar{\epsilon}\right)^{2j} \left(1-x-y\right)^{N-2j} \\
&\leq \lim_{N\to\infty} \sum_{j=0}^{N} \binom{N}{j} \left(x+y-\bar{\epsilon}\right)^{j} \left(1-x-y\right)^{N-j} \\
&= \lim_{N\to\infty} \left(1-\bar{\epsilon}\right)^N = 0,
\end{aligned}$$

which implies also $\lim_{N\to\infty} \mathbb{P}\left[T > \bar{t}\right] = 0$.

Case $x = y$. To study this case, let $y = x + \xi$ and let $\xi \to 0$. In this case, the probability of the decision time exceeding $\bar{t}$ becomes

$$f(x, N, \xi) = \mathbb{P}\left[T > \bar{t}\right] = \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{2j}\binom{2j}{j} x^j (x+\xi)^j \left(1-2x-\xi\right)^{N-2j}.$$

Consider $\lim_{\xi\to 0} f(x, N, \xi)$. We have that

$$\begin{aligned}
\lim_{\xi\to 0} f(x, N, \xi) &= \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{2j}\binom{2j}{j} x^{2j} \left(1-2x\right)^{N-2j} \\
&< \sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{2j} 2^{2j} x^{2j} \left(1-2x\right)^{N-2j} < 1,
\end{aligned}$$

where the first inequality follows from $\binom{2j}{j} < \sum_{l=0}^{2j} \binom{2j}{l} = 2^{2j}$, and the second inequality follows from

$$\sum_{j=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{2j} (2x)^{2j} \left(1-2x\right)^{N-2j} < \sum_{2j=0}^{N} \binom{N}{2j} (2x)^{2j} \left(1-2x\right)^{N-2j} = 1.$$

So $\lim_{\xi\to 0} f(x,N,\xi)$ exists, and since we know that also $\lim_{N\to\infty} f(x,N,\xi)$ exists, the limits are exchangeable in $\lim_{N\to\infty}\lim_{\xi\to 0} f(x,N,\xi)$ and

$$\lim_{N\to\infty}\lim_{\xi\to 0} f(x,N,\epsilon) = \lim_{\xi\to 0}\lim_{N\to\infty} f(x,N,\xi) = 0.$$

This concludes the proof of equation (2.21).

Assume now that $p_{1|1}(\bar{t}) > p_{0|1}(\bar{t})$. We distinguish between the case where $p_{1|1}(\bar{t}) > \frac{1}{2}$ and the case where $p_{0|1}(\bar{t}) < p_{1|1}(\bar{t}) \leq \frac{1}{2}$.

If $p_{1|1}(\bar{t}) > \frac{1}{2}$, then Lemma B.0.1 implies

$$\lim_{N\to\infty} \sum_{j=\lceil N/2\rceil}^{N} \binom{N}{j} p_{1|1}^{j}(\bar{t}) \left(1 - p_{1|1}(\bar{t})\right)^{N-j} = 1,$$

and, since $\lim_{N\to\infty}\beta_{1|1}(\bar{t},0,0) > \lim_{N\to\infty}\sum_{j=\lceil N/2\rceil}^{N}\binom{N}{j}p_{1|1}^{j}(\bar{t})\left(1-p_{1|1}(\bar{t})\right)^{N-j}$, we have also that $\lim_{N\to\infty}\beta_{1|1}(\bar{t},0,0) = 1$.

The case $p_{0|1}(\bar{t}) < p_{1|1}(\bar{t}) < \frac{1}{2}$ is more involved. We will see that in this case $\lim_{N\to\infty}\beta_{0|1}(\bar{t},0,0) = 0$. We start by observing that, from Lemma B.0.1,

$$\lim_{N\to\infty} \sum_{j=\lceil \frac{N}{2}\rceil}^{N} \binom{N}{j} p_{1|1}^{j}(\bar{t}) \left([1 - p_{1|1}(\bar{t})\right)^{N-j} = 0,$$

and in turn

$$\lim_{N\to\infty}\beta_{1|1}(\bar{t},0,0) = \lim_{N\to\infty}\sum_{j=1}^{\lfloor \frac{N}{2}\rfloor}\binom{N}{j}p_{1|1}^{j}(\bar{t}) \times \left(\sum_{i=0}^{j-1}\binom{N-j}{i}p_{0|1}^{i}(\bar{t})\left[1-p_{1|1}(\bar{t})-p_{0|1}(\bar{t})\right]^{N-j-i}\right).$$

The above expression can be written as follows

$$
\begin{aligned}
\lim_{N\to\infty} \beta_{1|1}(\bar{t},0,0) &= \lim_{N\to\infty} \sum_{h=1}^{N-2}\Bigg(\sum_{j=\lfloor \frac{h}{2} \rfloor+1}^{h} \binom{N}{j}\binom{N-j}{h-j} p_{0|1}^{h-j}(\bar{t}) p_{1|1}^{j}(\bar{t})\Bigg) \\
&\quad \times \Big(1-\big(p_{0|1}(\bar{t})p_{1|1}(\bar{t})\big)\Big)^{N-h} = \lim_{N\to\infty} \sum_{h=1}^{N-2} \binom{N}{h} \sum_{j=\lfloor \frac{h}{2} \rfloor+1}^{h} \binom{h}{j} p_{1|1}^{h-j}(\bar{t}) p_{0|1}^{j}(\bar{t}) \\
&\quad \times \Big(1-p_{1|1}(\bar{t})-p_{0|1}(\bar{t})\Big)^{N-h}
\end{aligned}
$$

where, for obtaining the second equality we used the fact $\binom{N}{j}\binom{N-j}{h-j} = \binom{N}{h}\binom{h}{j}$.

Similarly,

$$
\begin{aligned}
&\lim_{N\to\infty} \beta_{0|1}(\bar{t},0,0) \\
&= \lim_{N\to\infty} \sum_{h=1}^{N-2} \binom{N}{h} \sum_{j=\lfloor \frac{h}{2} \rfloor+1}^{h} \binom{h}{j} p_{0|1}^{h-j}(\bar{t}) p_{1|1}^{j}(\bar{t}) \Big(1-p_{1|1}(\bar{t})-p_{0|1}(\bar{t})\Big)^{N-h}.
\end{aligned}
$$

We prove now that $\lim_{N\to\infty} \beta_{0|1}(\bar{t},0,0) = 0$. To do so we will show that there exists $\bar{\epsilon}$ depending only on $p_{0|1}(\bar{t})$ and $p_{1|1}(\bar{t})$ such that

$$
\sum_{j=\lfloor \frac{h}{2} \rfloor+1}^{h} \binom{h}{j} p_{0|1}^{h-j}(\bar{t}) p_{1|1}^{j}(\bar{t}) < \Big(p_{0|1}(\bar{t})+p_{1|1}(\bar{t})-\bar{\epsilon}\Big)^{h}.
$$

To do so, let

$$
\epsilon(h) = p_{0|1}(\bar{t})+p_{1|1}(\bar{t}) - \sqrt[h]{\sum_{j=\lfloor \frac{h}{2} \rfloor+1}^{h} \binom{h}{j} p_{0|1}^{h-j}(\bar{t}) p_{1|1}^{j}(\bar{t})}.
$$

Because h is bounded, one can see that $\epsilon(h) > 0$ as the sum inside the root is always smaller than $(p_{0|1}(\bar{t}) + p_{1|1}(\bar{t}))^h$. Also

$$\begin{aligned}
\lim_{h\to\infty} \epsilon(h) &= \left(p_{0|1}(\bar{t}) + p_{1|1}(\bar{t})\right)\left[1 - \frac{\sqrt[h]{\sum_{j=\lfloor \frac{h}{2}\rfloor+1}^{h} \binom{h}{j} p_{0|1}^{h-j}(\bar{t}) p_{1|1}^{j}(\bar{t})}}{p_{0|1}(\bar{t}) + p_{1|1}(\bar{t})}\right] \\
&= \left(p_{0|1}(\bar{t}) + p_{1|1}(\bar{t})\right)\left[1 - \sqrt[h]{\frac{\sum_{j=\lfloor \frac{h}{2}\rfloor+1}^{h} \binom{h}{j} p_{0|1}^{h-j}(\bar{t}) p_{1|1}^{j}(\bar{t})}{\left(p_{0|1}(\bar{t}) + p_{1|1}(\bar{t})\right)^h}}\right] \\
&= p_{0|1}(\bar{t}) + p_{1|1}(\bar{t}),
\end{aligned}$$

as by Lemma B.0.1,

$$\lim_{h\to\infty} \frac{\sum_{j=\lfloor \frac{h}{2}\rfloor}^{h} \binom{h}{j} p_{0|1}^{h-j}(\bar{t}) p_{1|1}^{j}(\bar{t})}{\left(p_{0|1}(\bar{t}) + p_{1|1}(\bar{t})\right)^h} = 0.$$

Since by assumption, $p_{0|1}(\bar{t}) + p_{1|1}(\bar{t}) > 0$, we have that $\inf_{h\in\mathbb{N}} \epsilon(h) > 0$. By letting $\bar{\epsilon} := \inf_{h\in\mathbb{N}} \epsilon(h)$, we conclude that

$$\begin{aligned}
\lim_{N\to\infty} \beta_{0|1}(\bar{t},0,0) &\leq \sum_{h=1}^{N-2} \binom{N}{h} \left(p_{1|1}(\bar{t}) + p_{0|1}(\bar{t}) - \bar{\epsilon}\right)\left(1 - p_{1|1}(\bar{t}) - p_{0|1}(\bar{t})\right)^{N-h} \\
&\leq \sum_{h=0}^{N} \binom{N}{h} \left(p_{1|1}(\bar{t}) + p_{0|1}(\bar{t}) - \bar{\epsilon}\right)\left(1 - p_{1|1}(\bar{t}) - p_{0|1}(\bar{t})\right)^{N-h} \\
&= (1-\bar{\epsilon})^N = 0.
\end{aligned}$$

This concludes the proof.

Remark 2.5.2 *The earliest possible decision time $\bar{t}$ defined in* (2.17) *is the best performance that the fastest rule can achieve in terms of number of iterations required to provide the final decision.* □

2.5.1 The majority rule for varying values of N

We consider now the *majority* rule, i.e., the counting rule with threshold $q = \lfloor N/2 \rfloor + 1$. We start with the following result about the accuracy. Recall that $p_{\mathrm{w}|1}$ is the probability of wrong decision by a single SDM and that $p_{\mathrm{w}|1}^{(\mathrm{m})}(N)$ is the probability of wrong decision by a group of N SDMs implementing the majority rule (assuming H_1 is the correct hypothesis).

Proposition 2.5.3 (Accuracy under the majority rule) *Consider the q out of N SDA algorithm under Assumption 2.5. Assume $q = \lfloor N/2 \rfloor + 1$, i.e., the* majority *rule is adopted. Without loss of generality, assume H_1 is the correct hypothesis. Then the probability of error satisfies*

$$p_{\mathrm{w}|1}^{(\mathrm{m})}(N) = \sum_{j=\lfloor N/2 \rfloor+1}^{N} \binom{N}{j} p_{\mathrm{w}|1}^{j} \left(1 - p_{\mathrm{w}|1}\right)^{N-j}. \tag{2.23}$$

According to (2.23)*, the following characterization follows:*

1. *if $0 \leq p_{\mathrm{w}|1} < 1/2$, then $p_{\mathrm{w}|1}^{(\mathrm{m})}(N)$ is a monotonic decreasing function of N that approaches 0 asymptotically, that is,*

$$p_{\mathrm{w}|1}^{(\mathrm{m})}(N) > p_{\mathrm{w}|1}^{(\mathrm{m})}(N+2) \quad \textit{and} \quad \lim_{N\to\infty} p_{\mathrm{w}|1}^{(\mathrm{m})}(N) = 0;$$

2. *if $1/2 < p_{\mathrm{w}|1} \leq 1$, then $p_{\mathrm{w}|1}^{(\mathrm{m})}(N)$ is a monotonic increasing function of N that approaches 1 asymptotically, that is,*

$$p_{\mathrm{w}|1}^{(\mathrm{m})}(N) < p_{\mathrm{w}|1}^{(\mathrm{m})}(N+2) \quad \textit{and} \quad \lim_{N\to\infty} p_{\mathrm{w}|1}^{(\mathrm{m})}(N) = 1;$$

3. *if* $p_{\mathrm{w}|1} = 1/2$, *then* $p^{(\mathrm{m})}_{\mathrm{w}|1}(N) = 1/2$;

4. *if* $p_{\mathrm{w}|1} < 1/4$, *then*

$$p^{(\mathrm{m})}_{\mathrm{w}|1}(N) = \binom{N}{\lceil \frac{N}{2} \rceil} p_{\mathrm{w}|1}^{\lceil \frac{N}{2} \rceil} + o\left(p_{\mathrm{w}|1}^{\lceil \frac{N}{2} \rceil}\right) = \sqrt{N/(2\pi)}\,(4p_{\mathrm{w}|1})^{\lceil \frac{N}{2} \rceil} + o\left((4p_{\mathrm{w}|1})^{\lceil \frac{N}{2} \rceil}\right). \tag{2.24}$$

Proof: We start by observing that

$$\sum_{s=1}^{t} p_{0|1}(s; N, q = \lfloor N/2 \rfloor + 1) = \sum_{j=\lfloor N/2 \rfloor + 1}^{N} \binom{N}{j} \pi_{0|1}(t)^j \left(1 - \pi_{0|1}(t)\right)^{N-j}.$$

Since $p^{(\mathrm{m})}_{\mathrm{w}|1}(N) = \sum_{s=1}^{\infty} p_{0|1}(s; N, q = \lfloor N/2 \rfloor + 1)$, taking the limit for $t \to \infty$ in the above expression leads to

$$p^{(\mathrm{m})}_{\mathrm{w}|1}(N) = \sum_{j=\lceil \frac{N}{2} \rceil}^{N} \binom{N}{j} p_{\mathrm{w}|1}^{j} \left(1 - p_{\mathrm{w}|1}\right)^{N-j}.$$

Facts (i), (ii), (iii) follow directly from Lemma B.0.1 in Appendix B applied to equation (2.23). Equation (2.24) is a consequence of the Taylor expansion of (2.23):

$$\begin{aligned}\sum_{j=\lceil \frac{N}{2} \rceil}^{N} \binom{N}{j} p_{\mathrm{w}|1}^{j} (1 - p_{\mathrm{w}|1})^{N-j} &= \sum_{j=\lceil \frac{N}{2} \rceil}^{N} \binom{N}{j} p_{\mathrm{w}|1}^{j} (1 - (N-j)p_{\mathrm{w}|1} + o(p_{\mathrm{w}|1})) \\ &= \binom{N}{\lceil \frac{N}{2} \rceil} p_{\mathrm{w}|1}^{\lceil \frac{N}{2} \rceil} + o\left(p_{\mathrm{w}|1}^{\lceil \frac{N}{2} \rceil + 1}\right).\end{aligned}$$

Finally, Stirling's Formula implies $\lim_{N\to\infty} \binom{N}{\lceil \frac{N}{2} \rceil} = \sqrt{2N/\pi}\, 2^N$ and, in turn, the final expansion follows from $2^N = 4^{\lceil N/2 \rceil}/2$.

We discuss now the expected time required by the collective SDA algorithm to provide a decision when the *majority* rule is adopted. Our analysis is based again on Assumption 2.5 and on the assumption that H_1 is the correct hypothesis. We distinguish four cases based on different properties that the probabilities of wrong and correct decision of the single SDM might have:

(A1) the probability of correct decision is greater than the probability of wrong decision, i.e., $p_{\mathrm{c}|1} > p_{\mathrm{w}|1}$;

(A2) the probability of correct decision is equal to the probability of wrong decision, i.e., $p_{\mathrm{c}|1} = p_{\mathrm{w}|1} = 1/2$ and there exist t_0 and t_1 such that $\pi_{0|1}(t_0) = 1/2$ and $\pi_{1|1}(t_1) = 1/2$;

(A3) the probability of correct decision is equal to the probability of wrong decision, i.e., $p_{\mathrm{c}|1} = p_{\mathrm{w}|1} = 1/2$ and there exists t_1 such that $\pi_{1|1}(t_1) = 1/2$, while $\pi_{0|1}(t) < 1/2$ for all $t \in \mathbb{N}$ and $\lim_{t\to\infty} \pi_{0|1}(t) = 1/2$;

(A4) the probability of correct decision is equal to the probability of wrong decision, i.e., $p_{\mathrm{c}|1} = p_{\mathrm{w}|1} = 1/2$, and $\pi_{0|1}(t) < 1/2$, $\pi_{1|1}(t) < 1/2$ for all $t \in \mathbb{N}$ and $\lim_{t\to\infty} \pi_{0|1} = \lim_{t\to\infty} \pi_{1|1}(t) = 1/2$.

Note that, since Assumption 2.5 implies $p_{\mathrm{c}|1} + p_{\mathrm{w}|1} = 1$, the probability of correct decision in case (A1) satisfies $p_{\mathrm{c}|1} > 1/2$. Hence, in case (A1) and under

Assumption 2.5, we define $t_{<\frac{1}{2}} := \max\{t \in \mathbb{N} \mid \pi_{1|1}(t) < 1/2\}$ and $t_{>\frac{1}{2}} := \min\{t \in \mathbb{N} \mid \pi_{1|1}(t) > 1/2\}$.

Proposition 2.5.4 (Expected time under the *majority* rule) *Consider the q out of N SDA algorithm under Assumption 2.5. Assume $q = \lfloor N/2 \rfloor + 1$, that is, adopt the* majority *rule. Without loss of generality, assume H_1 is the correct hypothesis. Define the SDM properties (A1)-(A4) and the decision times t_0, t_1, $t_{<\frac{1}{2}}$ and $t_{>\frac{1}{2}}$ as above. Then the expected decision time satisfies*

$$\lim_{N\to\infty} \mathbb{E}\big[T|H_1, N, q = \lceil N/2 \rceil\big] = \begin{cases} \dfrac{t_{<\frac{1}{2}} + t_{>\frac{1}{2}} + 1}{2}, & \textit{if SDM satisfies (A1)}, \\[2ex] \dfrac{t_1 + t_0}{2}, & \textit{if SDM satisfies (A2)}, \\[2ex] +\infty, & \textit{if SDM satisfies (A3) or (A4)}. \end{cases}$$

Proof: We start by proving the equality for case (A1). Since, in this case we are assuming $p_{\mathrm{c}|1} > p_{\mathrm{w}|1}$, the definitions of $t_{<\frac{1}{2}}$ and $t_{>\frac{1}{2}}$ implies that $\pi_{1|1}(t) = 1/2$ for all $t_{<\frac{1}{2}} < t < t_{>\frac{1}{2}}$. Observe that

$$\sum_{s=1}^{t} p_{1|1}(s; N, q = \lfloor N/2 \rfloor + 1) = \sum_{h=\lfloor \frac{N}{2} \rfloor}^{N} \binom{N}{h} \pi_{1|1}^{h}(t) \Big(1 - \pi_{1|1}(t)\Big)^{N-h}.$$

Hence Lemma B.0.1 implies

$$\lim_{N\to\infty}\sum_{s=1}^{t} p_{1|1}(s;N,q=\lfloor N/2\rfloor+1)=\begin{cases}0, & \text{if } \; t\le t_{<\frac{1}{2}},\\ 1, & \text{if } \; t\ge t_{>\frac{1}{2}},\\ \frac{1}{2}, & \text{if } t_{<\frac{1}{2}}<t<t_{>\frac{1}{2}}\ ,\end{cases}$$

and, in turn, that

$$\lim_{N\to\infty} p_{1|1}(t;N,q=\lfloor N/2\rfloor+1)=\begin{cases}1/2, & \text{if } \; t=t_{<\frac{1}{2}}+1 \;\text{ and }\; t=t_{>\frac{1}{2}},\\ 0, & \text{otherwise.}\end{cases}$$

It follows

$$\begin{aligned}&\lim_{N\to\infty}\mathbb{E}\left[T|H_1,N,q=\lfloor N/2\rfloor+1\right]\\ &=\lim_{N\to\infty}\sum_{t=0}^{\infty} t\left(p_{0|1}(t;N,q=\lfloor N/2\rfloor+1)+p_{1|1}(t;N,q=\lfloor N/2\rfloor+1)\right)\\ &=\frac{1}{2}\left(t_{<\frac{1}{2}}+1+t_{>\frac{1}{2}}\right).\end{aligned}$$

This concludes the proof of the equality for case (A1).

We consider now the case (A2). Reasoning similarly to the previous case we have that

$$\lim_{N\to\infty} p_{1|1}(t_1;N,q=\lfloor N/2\rfloor+1)=1/2\text{and}\lim_{N\to\infty} p_{0|1}(t_0;N,q=\lfloor N/2\rfloor+1)=1/2,$$

from which it easily follows that $\lim_{N\to\infty}\mathbb{E}\left[T|H_1,N,q=\lfloor N/2\rfloor+1\right]=\frac{1}{2}\left(t_0+t_1\right)$.

For case (A3), it suffices to note the following implication of Lemma B.0.1: if, for a given $i \in \{0,1\}$, we have $\pi_{i|1}(t) < 1/2$ for all $t \in \mathbb{N}$, then $\lim_{N\to\infty} p_{i|1}(t; N, q = \lfloor N/2 \rfloor + 1) = 0$ for all $t \in \mathbb{N}$. The analysis of the case (A4) is analogous to that of case (A3).

Remark 2.5.5 *The cases where $p_{\mathrm{w}|1} > p_{\mathrm{c}|1}$ and where there exists t_0 such that $\pi_{0|1}(t_0) = 1/2$ while $\pi_{1|1}(t) < 1/2$ for all $t \in \mathbb{N}$ and $\lim_{t\to\infty} \pi_{1|1}(t) = 1/2$, can be analyzed similarly to the cases (A1) and (A3). Moreover, the most recurrent situation in applications is the one where there exists a time instant t such that $\pi_{1|1}(t) < 1/2$ and $\pi_{1|1}(t+1) > 1/2$, which is equivalent to the above case (A1) with $t_{>\frac{1}{2}} = t_{<\frac{1}{2}}+1$. In this situation we trivially have $\lim_{N\to\infty} \mathbb{E}\left[T|H_1, N, q = \lceil N/2 \rceil\right] = t_{>\frac{1}{2}}$.* □

2.5.2 Fixed N and varying q

We start with a simple result characterizing the expected decision time.

Proposition 2.5.6 *Given a group of N SDMs running the q* out of *N* SDA*, for $j \in \{0,1\}$,*

$$\mathbb{E}[T|H_j, N, q = 1] \leq \mathbb{E}[T|H_j, N, q = 2] \leq \cdots \leq \mathbb{E}[T|H_j, N, q = N].$$

The above proposition states that the expected number of iterations required to provide a decision constitutes a nondecreasing sequence for increasing value of

q. Similar monotonicity results hold true also for $p_{\mathrm{c}|j}(N,q)$, $p_{\mathrm{w}|j}(N,q)$, $p_{\mathrm{nd}|j}(N,q)$ even though restricted only to $\lfloor N/2 \rfloor + 1 \leq q \leq N$.

Proposition 2.5.7 *Given a group of N SDMs running the q* out of *N* SDA*, for $j \in \{0,1\}$,*

$$p_{\mathrm{c}|j}(N, q = \lfloor N/2 \rfloor + 1) \geq p_{\mathrm{c}|j}(N, q = \lfloor N/2 \rfloor + 2) \geq \cdots \geq p_{\mathrm{c}|j}(N, q = N),$$

$$p_{\mathrm{w}|j}(N, q = \lfloor N/2 \rfloor + 1) \geq p_{\mathrm{w}|j}(N, q = \lfloor N/2 \rfloor + 2) \geq \cdots \geq p_{\mathrm{w}|j}(N, q = N),$$

$$p_{\mathrm{nd}|j}(N, q = \lfloor N/2 \rfloor + 1) \leq p_{\mathrm{nd}|j}(N, q = \lfloor N/2 \rfloor + 2) \leq \cdots \leq p_{\mathrm{nd}|j}(N, q = N).$$

We believe that similar monotonic results hold true also for $1 \leq q \leq \lfloor N/2 \rfloor$. In particular, here is our conjecture: if N is odd, the single SDM has the *almost-sure* decision and the single SDM is more likely to provide the correct decision than the wrong decision, that is, $p_{\mathrm{c}|j} + p_{\mathrm{w}|j} = 1$ and $p_{\mathrm{c}|j} > p_{\mathrm{w}|j}$, then

$$p_{\mathrm{c}|j}(N, q = 1) \leq p_{\mathrm{c}|j}(N, q = 2) \leq \cdots \leq p_{\mathrm{c}|j}(N, q = \lfloor N/2 \rfloor + 1),$$

$$p_{\mathrm{w}|j}(N, q = 1) \geq p_{\mathrm{w}|j}(N, q = 2) \geq \cdots \geq p_{\mathrm{w}|j}(N, q = \lfloor N/2 \rfloor + 1).$$

These chains of inequalities are numerically verified in some examples in Section 2.6.

2.6 Numerical analysis

The goal of this section is to numerically analyze the models and methods described in previous sections. In all the examples, we assume that the sequential binary test run by each SDMs is the classical sequential probability ratio test (SPRT) developed in 1943 by Abraham Wald. To fix some notation, we start by briefly reviewing the SPRT. Let X be a random variable with distribution $f(x;\theta)$ and assume the goal is to test the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1$. For $i \in \{1, \dots, N\}$, the i-th SDM takes the observations $x_i(1), x_i(2), x(3), \dots$, which are assumed to be independent of each other and from the observations taken by all the other SDMs. The log-likelihood ratio associated to the observation $x_i(t)$ is

$$\lambda_i(t) = \log \frac{f(x_i(t), \theta_1)}{f(x_i(t), \theta_0)}. \tag{2.25}$$

Accordingly, let $\Lambda_i(t) = \sum_{h=1}^{t} \lambda_i(h)$ denote the sum of the log-likelihoods up to time instant t. The i-th SDM continues to sample as long as $\eta_0 < \Lambda_i(t) < \eta_1$, where η_0 and η_1 are two pre-assigned thresholds; instead sampling is stopped the first time this inequality is violated. If $\Lambda_i(t) < \eta_0$, then the i-th SDM decides for $\theta = \theta_0$. If $\Lambda_i(t) > \eta_1$, then the i-th SDM decides for $\theta = \theta_1$.

To guarantee the *homogeneity property* we assume that all the SDMs have the same thresholds η_0 and η_1. The threshold values are related to the accuracy of the SPRT as described in the classic Wald's method [3]. We shortly review this method next. Assume that, for the single SDM, we want to set the thresholds η_0 and η_1 in such a way that the probabilities of misdetection (saying H_0 when H_1 is correct, i.e., $\mathbb{P}[\text{say } H_0|H_1]$) and of false alarm (saying H_1 when H_0 is correct, i.e., $\mathbb{P}[\text{say } H_1|H_0]$) are equal to some pre-assigned values $p_{\text{misdetection}}$ and $p_{\text{false alarm}}$. Wald proved that the inequalities $\mathbb{P}[\text{say } H_0 \,|\, H_1] \leq p_{\text{misdetection}}$ and $\mathbb{P}[\text{say } H_1 \,|\, H_0] \leq p_{\text{false alarm}}$ are achieved when η_0 and η_1 satisfy $\eta_0 \leq \log \frac{p_{\text{misdetection}}}{1-p_{\text{false alarm}}}$ and $\eta_1 \geq \log \frac{1-p_{\text{misdetection}}}{p_{\text{false alarm}}}$. As customary, we adopt the equality sign in these inequalities for the design of η_0 and η_1. Specifically, in all our examples we assume that $p_{\text{misdetection}} = p_{\text{false alarm}} = 0.1$ and, in turn, that $\eta_1 = -\eta_0 = \log 9$.

We provide numerical results for observations described by both discrete and continuous random variables. In case X is a discrete random variable, we assume that $f(x;\theta)$ is a binomial distribution

$$f(x;\theta) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x}, & \text{if } x \in \{0,1,\dots,n\}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.26}$$

where n is a positive integer number. In case X is a continuous random variable, we assume that $f(x;\theta)$ is a Gaussian distribution with mean θ and variance σ^2

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\theta)^2/2\sigma^2}. \tag{2.27}$$

The key ingredient required for the applicability of Propositions 2.3.1 and 2.3.2

is the knowledge of the probabilities $\{p_{0|0}(t), p_{1|0}(t)\}_{t\in\mathbb{N}}$ and $\{p_{0|1}(t), p_{1|1}(t)\}_{t\in\mathbb{N}}$. Given thresholds η_0 and η_1, then probabilities can be computed according to the method described in the Appendix C (respectively Appendix C) for X discrete (respectively X continuous) random variable.

We provide three sets of numerical results. Specifically, in Example 2.6.1 we emphasize the tradeoff between accuracy and expected decision time as a function of the number of SDMs. In Example 2.6.2 we concentrate on the monotonic behaviors that the q *out of* N SDA algorithm exhibits both when N is fixed and q varies and when q is fixed and N varies. In Example 2.6.3 we compare the *fastest* and the *majority* rule. Finally, Section 2.6.1 discusses drawing connections between the observations in Example 2.6.3 and the cognitive psychology presentation introduced in Section 2.1.3.

Example 2.6.1 (Tradeoff between accuracy and expected decision time) *This example emphasizes the tradeoff between accuracy and expected decision time as a function of the number of SDMs. We do that for the* fastest *and the* majority *rules. We obtain our numerical results for odd sizes of group of SDMs ranging from* 1 *to* 61*. In all our numerical examples, we compute the values of the thresholds* η_0 *and* η_1 *according to Wald's method by posing* $p_{\text{misdetection}} = p_{\text{false alarm}} = 0.1$ *and, therefore,* $\eta_1 = \log 9$ *and* $\eta_0 = -\log 9$.

For a binomial distribution $f(x;\theta)$ *as in* (2.26), *we provide our numerical results under the following conditions: we set* $n = 5$*; we run our computations for three different pairs* (θ_0, θ_1)*; precisely we assume that* $\theta_0 = 0.5 - \epsilon$ *and* $\theta_1 = 0.5 + \epsilon$ *where* $\epsilon \in \{0.02, 0.05, 0.08\}$*; and* $H_1 : \theta = \theta_1$ *is always the correct hypothesis. For any pair* (θ_0, θ_1) *we perform the following three actions in order*

1. *we compute the probabilities* $\left\{p_{0|1}(t), p_{1|1}(t)\right\}_{t\in\mathbb{N}}$ *according to the method described in Appendix C;*

2. *we compute the probabilities* $\left\{p_{0|1}(t;N,q), p_{1|1}(t;N,q)\right\}_{t\in\mathbb{N}}$ *for* $q = 1$ *and* $q = \lfloor N/2 \rfloor + 1$ *according to the formulas reported in Proposition 2.3.1;*

3. *we compute probability of wrong decision and expected time for the group of SDMs exploiting the formulas*

$$p_{\mathrm{w}|1}(N,q) = \sum_{t=1}^{\infty} p_{0|1}(t;N,q);\ \mathbb{E}[T|H_1,N,q] = \sum_{t=1}^{\infty}(p_{0|1}(t;N,q) + p_{1|1}(t;N,q))t.$$

According to Remark 2.2.7, since we consider only odd numbers N *of SDMs, since* $q \leq \lceil N/2 \rceil$ *and since each SDM running the SPRT has the* almost-sure decisions *property, then* $p_{\mathrm{w}|1}(N,q) + p_{\mathrm{c}|1}(N,q) = 1$*. In other words, the probability of no-decision is equal to* 0 *and, hence, the accuracy of the SDA algorithms is characterized only by the probability of wrong decision and the probability of correct decision. In our analysis we select to compute the probability of wrong decision.*

For a Gaussian distribution $f(x;\theta,\sigma)$, we obtain our numerical results under the following conditions: the two hypothesis are $H_0 : \theta = 0$ and $H_1 : \theta = 1$; we run our computations for three different values of σ; precisely $\sigma \in \{0.5, 1, 2\}$; and $H_1 : \theta = 1$ is always the correct hypothesis. To obtain $p_{\mathrm{w}|1}(N,q)$ and $\mathbb{E}[T|H_1,N,q]$ for a given value of σ, we proceed similarly to the previous case with the only difference that $\left\{p_{0|1}(t), p_{1|1}(t)\right\}_{t\in\mathbb{N}}$ are computed according to the procedure described in Appendix C.

The results obtained for the fastest *rule are depicted in Figure 2.2, while the results obtained for the* majority *rule are reported in Figure 2.3.*

Some remarks are now in order. We start with the fastest *rule. A better understanding of the plots in Figure 2.2 can be gained by specifying the values of the earliest possible decision time $\bar{t}$ defined in (2.17) and of the probabilities $p_{1|1}(\bar{t})$ and $p_{0|1}(\bar{t})$. In our numerical analysis, for each pair (θ_0, θ_1) considered and for both discrete and continuous measurements X, we had $\bar{t} = 1$ and $p_{1|1}(\bar{t}) > p_{0|1}(\bar{t})$. As expected from Proposition 2.5.1, we can see that the* fastest *rule significantly reduces the expected number of iterations required to provide a decision. Indeed, as N increases, the expected decision time $\mathbb{E}[T|H_1, N, q = 1]$ tends to 1. Moreover, notice that $p^{(\mathrm{f})}_{\mathrm{w}|1}(N)$ approaches 0; this is in accordance with equation (2.18).*

As far as the majority *rule is concerned, the results established in Proposition 2.5.3 and in Proposition 2.5.4 are confirmed by the plots in Figure 2.3. In-*

deed, since for all the pairs (θ_0, θ_1) *we have considered, we had* $p_{\mathrm{w}|1} < 1/2$*, we can see that, as expected from Proposition 2.5.3, the probability of wrong decision goes to* 0 *exponentially fast and monotonically as a function of the size of the group of the SDMs. Regarding the expected time, in all the cases, the expected decision time* $\mathbb{E}[T|H_1, N, q = \lfloor N/2 \rfloor + 1]$ *quickly reaches a constant value. We numerically verified that these constant values corresponded to the values predicted by the results reported in Proposition 2.5.4.*

Example 2.6.2 (Monotonic behavior) *In this example, we analyze the performance of the general* q out of N *aggregation rule, as the number of SDMs* N *is varied, and as the aggregation rule itself is varied. We obtained our numerical results for odd values of* N *ranging from* 1 *to* 35 *and for values of* q *comprised between* 1 *and* $\lfloor N/2 \rfloor + 1$*. Again we set the thresholds* η_0 *and* η_1 *equal to* $\log(-9)$ *and* $\log 9$*, respectively. In this example we consider only the Gaussian distribution with* $\sigma = 1$*. The results obtained are depicted in Figure 2.4, where the following monotonic behaviors appear evident:*

1. *for fixed* N *and increasing* q*, both the probability of correct decision and the decision time increases;*

2. *for fixed* q *and increasing* N*, the probability of correct decision increases while the decision time decreases.*

The fact that the decision time increases for fixed N and increasing q has been established in Proposition 2.5.6. The fact that the probability of correct decision increases for fixed N and increasing q validates the conjecture formulated at the end of Section 2.5.2.

Example 2.6.3 (Fastest versus majority, at fixed group accuracy) *As we noted earlier, Figures 2.2-2.3 show that the* majority *rule increases remarkably the accuracy of the group, while the* fastest *rule decreases remarkably the expected number of iteration for the SDA to reach a decision. It is therefore reasonable to pose the following question: if the local accuracies of the SDMs were set so that the accuracy of the group is the same for both the* fastest *and the* majority *fusion rule, which of the two rules requires a smaller number of observations to give a decision. That is, at equal accuracy, which of the two rules is optimal as far as decision time is concerned.*

In order to answer this question, we use a bisection on the local SDM accuracies. We apply the numerical methods presented in Proposition 2.3.1 to find the proper local thresholds that set the accuracy of the group to the desired value $p_{\mathrm{w}|1}$. Different local accuracies are obtained for different fusion rules and this evaluation needs to be repeated for each group size N.

In these simulations, we assume the random variable X is Gaussian with vari-

ance $\sigma = 2$. The two hypotheses are $H_0 : \theta = 0$ and $H_1 : \theta = 1$. The numerical results are shown in Figure 2.5 and discussed below.

As is clear from the plots, the strategy that gives the fastest decision with the same accuracy varies with group size and desired accuracy. The left plot in Figure 2.5 illustrates that, for very high desired group accuracy, the majority *rule is always optimal. As the accuracy requirement is relaxed, the* fastest *rule becomes optimal for small groups. Moreover, the group size at which the switch between optimal rules happens, varies for different accuracies. For example, the middle and right plot in Figure 2.5 illustrate that while the switch happens at $N = 5$ for a group accuracy $p^{(\mathrm{m})}_{\mathrm{w}|1} = p^{(\mathrm{f})}_{\mathrm{w}|1} = 0.05$ and at $N = 9$ for $p^{(\mathrm{m})}_{\mathrm{w}|1} = p^{(\mathrm{f})}_{\mathrm{w}|1} = 0.1$.*

We summarize our observations about which rule is optimal (i.e., which rule requires the least number of observations) as follows:

1. *the optimal rule varies with the desired network accuracy, at fixed network size;*

2. *the optimal rule varies with the desired network size, at fixed network accuracy; and*

3. *the change in optimality occurs at different network sizes for different accuracies.*

2.6.1 Decision making in cognitive psychology revisited

In this section we highlight some interesting relationships between our results in sequential decision aggregation (SDA) and some recent observations about mental behavior from the cognitive psychology literature. Starting with the literature review in Subsection 2.1.3, our discussion here is based upon the following assumptions:

1. SDA models multi-modal integration in cognitive information processing (CIP),
2. the number of SDMs correspond to the number of sensory modalities in CIP,
3. the expected decision time in the SDA setup is analogous to the reaction time in CIP, and
4. the decision probability in the SDA setup is analogous to the firing rate of neurons in CIP.

Under these assumptions, we relate our SDA analysis to four recent observations reported in the CIP literature. In short, the *fastest* and *majority* rules appear to emulate behaviors that are similar to the ones manifested by the brain under various conditions. These correspondences are summarized in Table 2.1 and described in the following paragraphs.

First, we look at the observation in CIP that multi-modal sites can exhibit suppressive behaviors (first row in Table 2.1). We find that suppressive behavior is not contradictory with the nature of such a site. Indeed, Proposition 2.5.1 describes situations where an increased group size degrades the decision accuracy of a group using the *fastest* rule.

Second, we look at the observation in CIP that, for some high-intensity stimuli, the firing rate of multi-modal integration sites is similar to the firing rate of uni-modal integration sites (second row in Table 2.1). This similarity behavior appears related to behaviors observed in Figure 2.5. The second and third plots in Figure 2.5 illustrate how, in small groups with high individual accuracy and relatively low group decision accuracy, the *fastest* rule is optimal. Since a multi-model integration site implementing a fastest aggregation rule behaves similarly to a uni-modal integration site, our result give a possible optimality interpretation of the observed "multi-modal similar to uni-modal" behavior.

Third, we look at the observation in CIP that activation of multi-modal integration sites is often accompanied with an increase in the accuracy as compared to the accuracy of a uni-sensory integration site (third and forth rows in Table 2.1). The first plot in Figure 2.5 shows that when the required performance is a high accuracy, the majority rule is better than the fastest. Indeed Proposition 2.5.3 proves that, for the *majority* rule, the accuracy monotonically increases with the group size, sometimes exponentially.

Fourth, we look at the observation in CIP that, even under the same type of stimuli, the stimuli strength affects the additivity of the neuron firing, i.e., the suppressive, additive, sub-additive or super-additive behavior of the firing rates. Additionally, scientists have observed that depending on the intensity of the stimuli, various areas of the brain are activated when processing the same type of stimuli [12, 13, 11, 17]. A possible explanation for these two observed behaviors is that the brain processes information in a way that maintains optimality. Indeed, our comparison in the middle and right parts of Figure 2.5 shows how the fastest rule is optimal when individual SDMs are highly accurate (strong and intact stimuli) and, vice versa, the majority rule is optimal when individual SDMs are relatively inaccurate (weak and degraded stimuli).

We observed in the middle and right part of Figure 2.5 that, for high individual accuracies, the *fastest* rule is more efficient than the *majority* rule. We reach this conclusion by noting two observations: first, smaller group sizes require higher local accuracies than larger group sizes in order to maintain the same group accuracy; second, the *fastest* rule is optimal for small groups while the *majority* rule is always optimal for larger groups.

2.7 Conclusion

In this work, we presented a complete analysis of how a group of SDMs can collectively reach a decision about the correctness of a hypothesis. We presented a numerical method that made it possible to completely analyze and understand interesting fusion rules of the individuals decisions. The analysis we presented concentrated on two aggregation rules, but a similar analysis can be made to understand other rules of interest. An important question we were able to answer was the one relating the size of the group and the overall desired accuracy to the optimal decision rules. We were able to show that no single rule is optimal for all group sizes or for various desired group accuracy.

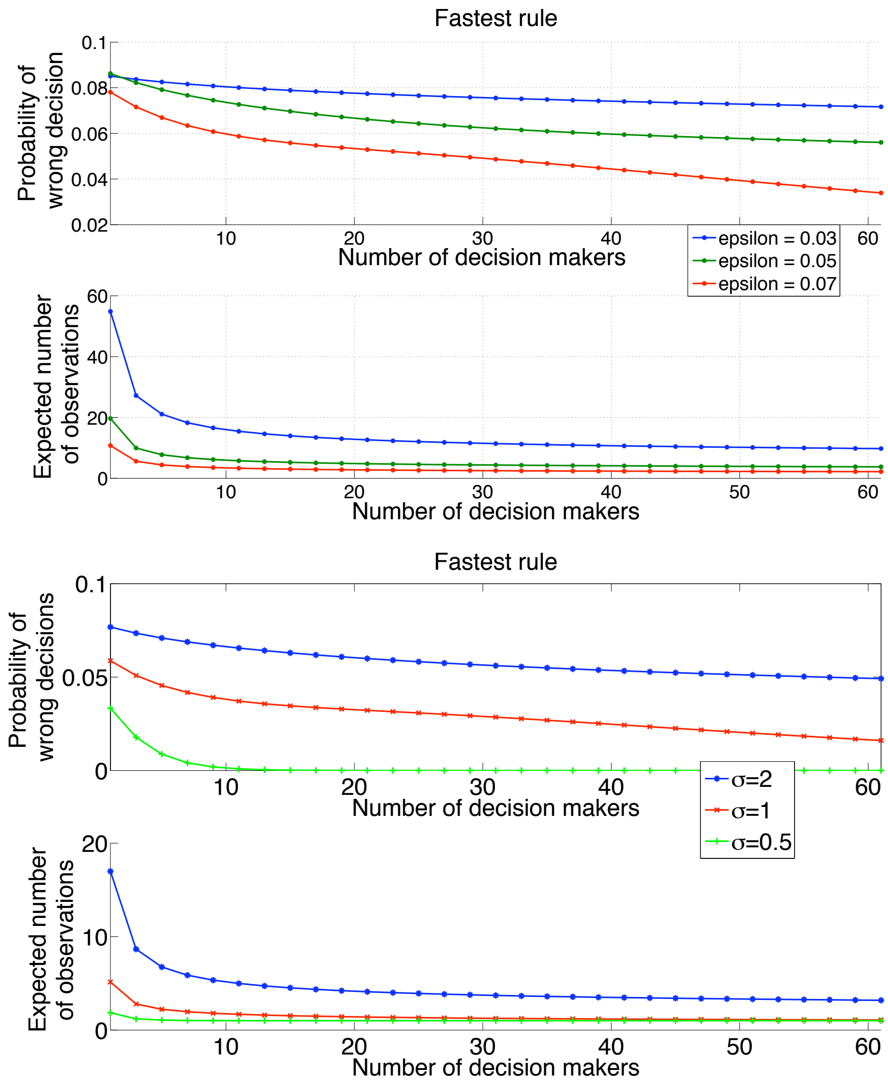


Figure 2.2: Behavior of the probability of wrong decision and of the expected number of iterations required to provide a decision as the number of SDMs increases when the *fastest* rule is adopted. In Figure (a) we consider the binomial distribution, in Figure (b) the Gaussian distribution.

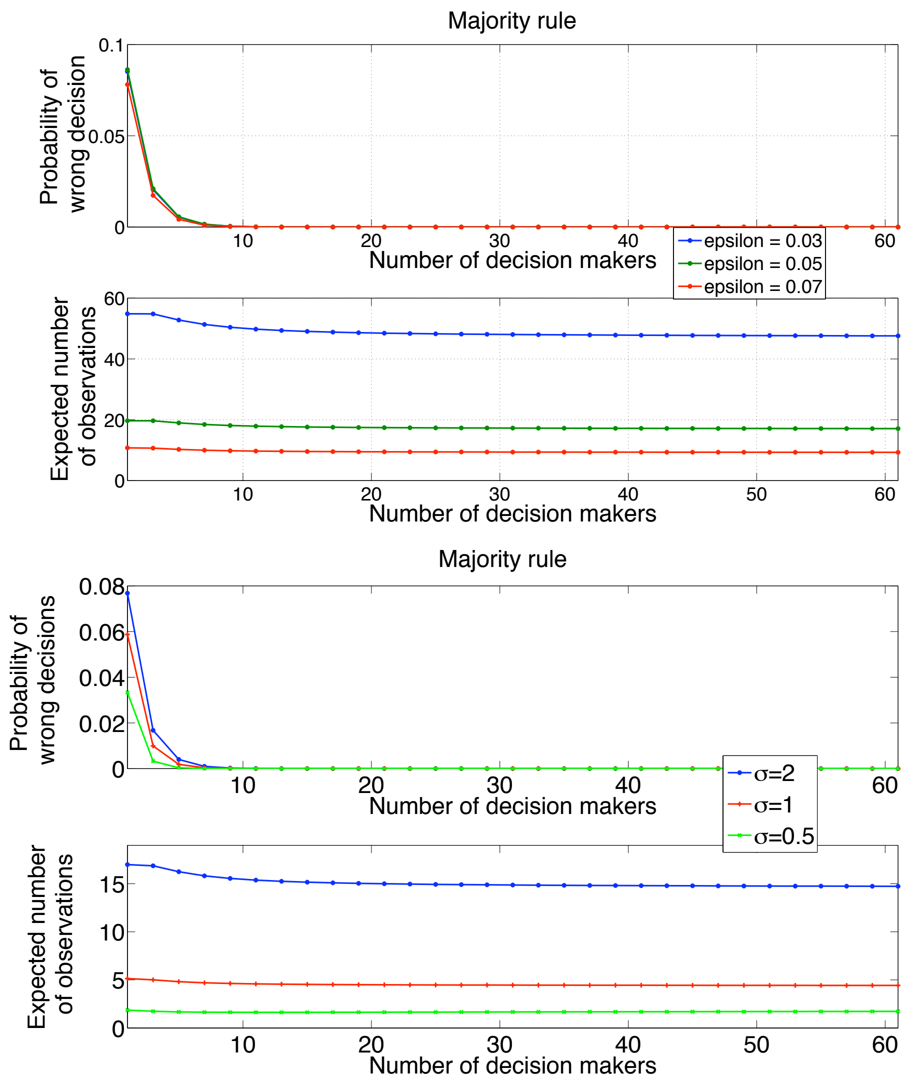


Figure 2.3: Behavior of the probability of wrong decision and of the expected number of iterations required to provide a decision as the number of SDMs increases when the *majority* rule is adopted. In Figure (a) we consider the binomial distribution, in Figure (b) the Gaussian distribution.

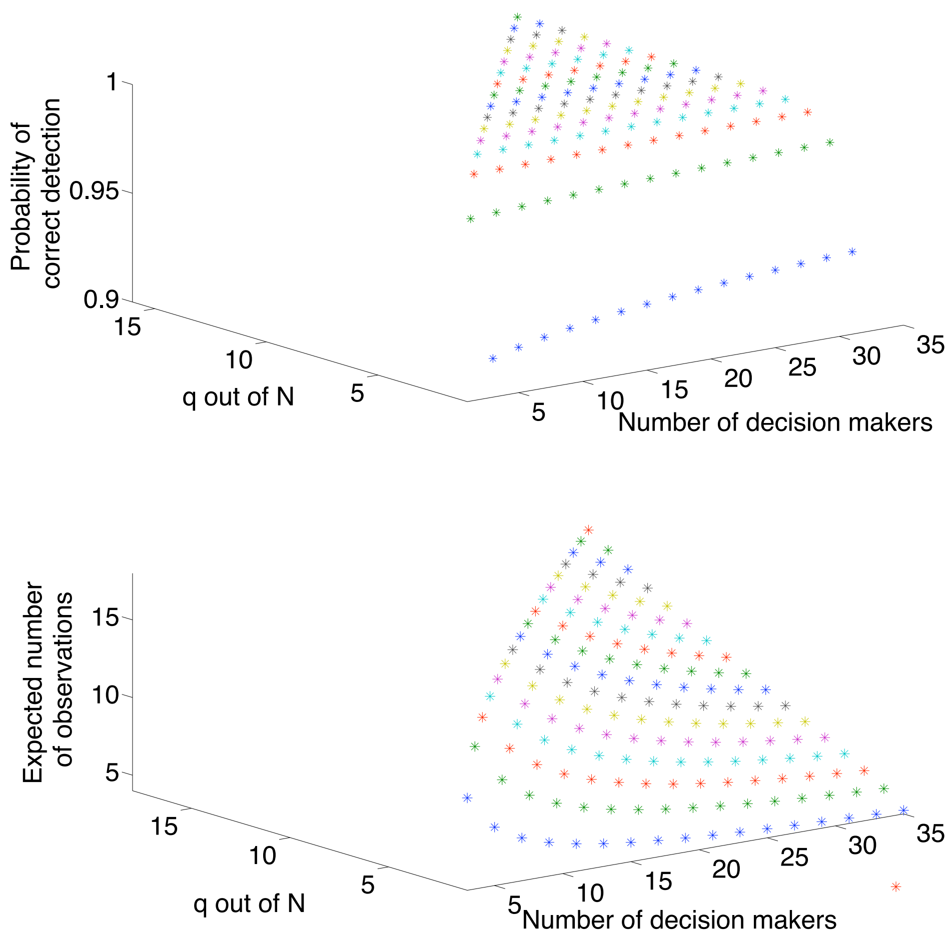


Figure 2.4: Probability of correct detection (left figure) and expected decision time (right figure) for the q *out of* N rule, plotted as a function of network size N and accuracy threshold q.

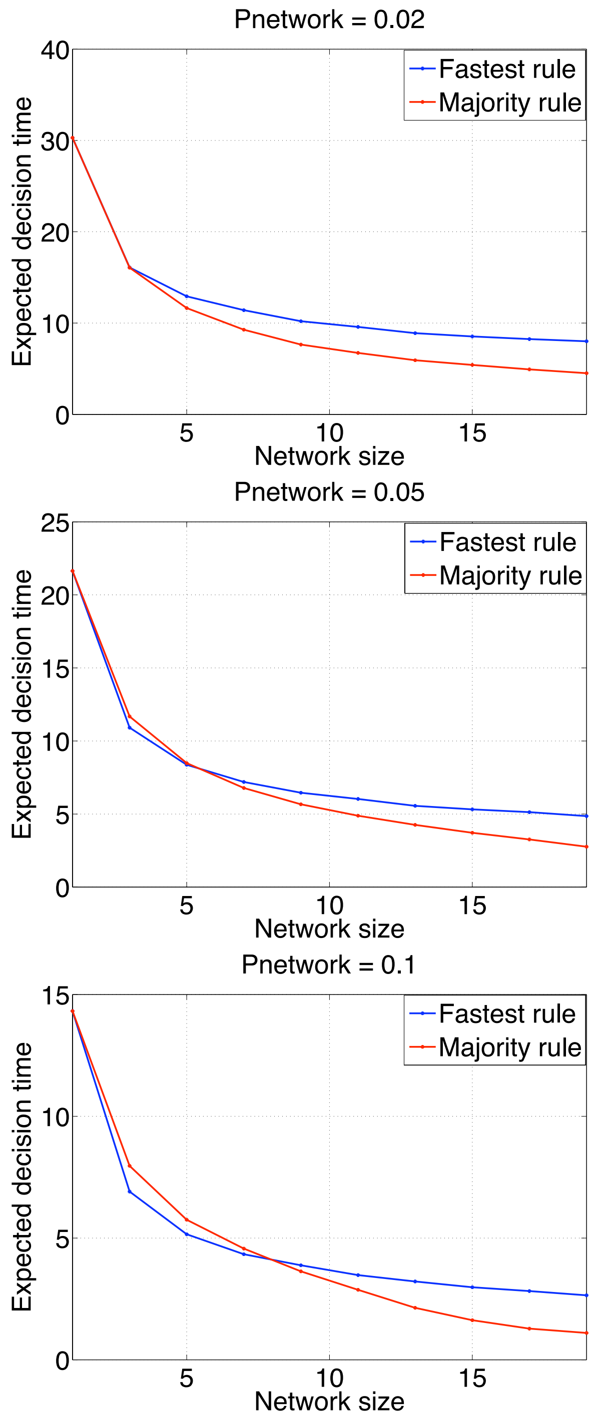


Figure 2.5: Expected decision time for the *fastest* and the *majority* rules versus group size N, for various network accuracy levels.

Table 2.1: Cognitive Psychology - Engineering dictionary of correspondences.

Multi-sensory integration sites	*Sequential decision aggregation*
Suppressive behavior of firing rate	Decision probability decreases with increasing N
Sub-additive behavior of firing rates	Probability of decision slightly increases with increasing N
Additive behavior of firing rates	Decision probability linearly increases with increasing N
Super-additive behavior of firing rates	Decision probability exponentially increases with increasing N

Chapter 3

Distributed Sequential Algorithms for Regional Source Localization

3.1 Introduction

We study in this chapter the problem of source localization, where a group of sensors sense, transmit and process information with the objective of confining the source location to a region inside a terrain of interest.

3.1.1 Problem description and motivation

Formally, the problem is the following: A source at an unknown location in a bounded region Q transmits a power signal. N sensors receive noisy and decayed versions of the signal, and they can communicate and exchange measurements. The environment Q is divided into M regions W_α, where $\alpha \in \{1, \ldots, M\}$. The objective of the sensors is to find which region contains the source.

We pose the problem as a multiple hypothesis testing problem, where hypothesis H_α is true if the source lies in the region W_α. We assume no prior knowledge about the location of the source and therefore model the source location as a uniformly distributed random variable over the environment Q; any prior information about the source location can be incorporated in the location density function. We adopt the log-normal fading model for the propagation of the received signal power. The noise added to the log of the power is Gaussian with zero mean and a known variance σ^2.

3.1.2 Chapter contributions

The contributions of this chapter are three-fold.

First, we formulate the source localization problem in a novel multi-hypothesis testing setting. We analyze properties of the Maximum A Posteriori (MAP) algorithm that requires the computation of a finite number of integrals which is to be compared to the need to solve a non-linear, non-convex problem in the classical source localization problem. We provide a proof of almost sure convergence of the MAP solution asymptotically in the limit of a large number of measurements, a step that tends to be missing in all of the work presented earlier in the source localization literature.

Second, inspired by the proof of convergence of the MAP solution, we propose and implement a distributed sequential regional localization algorithm: *Sense, Transmit & Test*. This algorithm allows for sequential sensing, transmission and testing at each processor. We allow each processor to have one or multiple regions of responsibility and relate the probability of error for each processor in the case of multiple regions to the probability of error in the case of a single region. We also show that the test ends in a finite time under mild conditions on the sensor locations.

Third, we illustrate the results of the *Sense, Transmit & Test* and show how the expected decision time for a network increases with the required accuracy and noise. We also provide numerical results illustrating how it is possible to increase the level of localization accuracy at the expense of the expected decision time for the network for a fixed decision accuracy.

3.1.3 Chapter organization

The chapter proceeds as follows: we formulate the problem as a multi-hypothesis testing problem in Section 3.2. We present a distributed algorithm to solve the problem in Section 3.3. We present in Section 3.4 numerical results showing the performance of the algorithm as various parameters are changed. We conclude in Section 3.5.

3.1.4 Preliminary concepts

We present here a few preliminary concepts that will be useful in this chapter.

Lemma 3.1.1 (On the Dirac delta function) *Let $\delta_{\mathrm{D}}(x)$ be the Dirac delta function. Given a scalar function $g : \mathbb{R} \to \mathbb{R}$, let S be the set of scalar z where $g(z) = 0$. If g is differentiable and $g'(z) \neq 0$ for $z \in S$, then*

$$\delta_{\mathrm{D}}(g(x)) = \sum_{z \in S} \frac{\delta_{\mathrm{D}}(x - z)}{|g'(z)|}. \tag{3.1}$$

Definition 3.1.2 (Partitions and Voronoi partitions) *A partition $\{W_1, \ldots, W_M\}$ of a space Q is a collection of closed subsets of Q with the following properties: each W_α has positive measure, each intersection $W_\alpha \cap W_\beta$ has zero measure, and $\cup_{\alpha=1}^{M} W_\alpha = Q$.*

Given distinct positions $\{p_1, \dots, p_N\} \in \mathbb{R}^2$*, the* Voronoi partition $\{V_1, \dots, V_N\}$ *of* Q *is defined by*

$$V_i = \{q \in Q \mid \|q - p_i\| \leq \|q - p_j\|, \forall j \neq i\}, \text{ for } i \in \{1, \dots, N\}.$$

3.2 Source localization as multi-hypothesis testing

We start this section by introducing the model and the problem definition.

3.2.1 Model and problem definition

Consider a compact connected environment $Q \subset \mathbb{R}^2$. Suppose that there are N sensors placed at positions $q_i \in Q$ with $i \in \{1, \dots, N\}$, and that the source located at an unknown location $s \in Q$ transmits a signal whose power undergoes lognormal shadowing summarized as follows. The average power loss for an arbitrary Transmitter-Receiver separation is expressed as a function of distance by using a path loss exponent $\rho > 2$. For reasons to be explained shortly, we work with a slight modification of the traditionally used model. The adopted model for the received power at a sensor i is $P_i = \frac{Pd_0}{d_0 + \|q_i - s\|^\rho}$, where ρ indicates the rate at which the power loss increases with distance. The nominal distance d_0

is chosen so that the received power in the vicinity of the source is almost equal to the transmitted power P at the source. Note that while this model gets rid of the singularity at the source, it converges to the same behavior as the classical model used in communication literature $P_i = \frac{P}{\|q_i - s\|^\rho}$, when the distance $\|q_i - s\|$ is large. Here P is the power received at a unit distance from the source. The received power becomes

$$\ln P_i = \ln(P d_0) - \ln(d_0 + \|q_i - s\|^\rho) + n_i, \tag{3.2}$$

where n_i is the noise associated with sensor i, and all n_i are independent and identically distributed (i.i.d) Gaussian random variables with zero mean and known variance σ^2. The joint probability density function of the received power $P_r = [P_1, \ldots, P_N]^T$, conditioned on the source location $y \in Q$ is

$$\mathbb{p}(P_1, \ldots, P_N | y) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\Big(- \frac{\sum_{i=1}^{N} \Big(\ln P_i - \ln(\frac{P d_0}{d_0 + \|q_i - y\|^\rho}) \Big)^2}{2\sigma^2} \Big). \tag{3.3}$$

Problem 3.2.1 (MAP point localization problem) *Compute the position that maximizes the conditional density of the joint observations, that is compute*

$$y^* = \operatorname*{argmax}_{y \in Q} \mathbb{p}(P_1, \ldots, P_N | y) \mathbb{P}(y).$$

Problem 3.2.1 is a nonlinear nonconvex optimization problem. Attempts to solve this problem, usually revert to relaxing the problem or approximating its solution without providing a convergence analysis. In this chapter we look for a regional localization, so the conditioning on the exact position y in (3.3) is replaced by a conditioning on the source being in a region W_i. The environment Q with area A is divided into M regions $\{W_1, \dots, W_M\}$ with positive areas $\{A_1, \dots, A_M\}$. The hypothesis H_α is true if and only if $s \in W_\alpha$.

Problem 3.2.2 (MAP regional localization problem) *Compute the hypothesis H_α that maximizes the posterior of the joint observations, that is, compute*

$$\alpha^* = \underset{\alpha \in \{1,\dots,M\}}{\operatorname{argmax}} \; \mathbb{p}(P_1, \dots, P_N | H_\alpha)\mathbb{P}(H_\alpha). \tag{3.4}$$

3.2.2 Regional posterior density

Assuming no prior knowledge about the location of the source, the density describing $s \in Q$ is

$$\mathbb{p}(s) = \begin{cases} 1/A, & \text{if } s \in Q, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 3.2.3 (Repeated measurements) *The ith sensor takes k repeated i.i.d. noisy measurements and computes the average of the logarithms of the measurements*

$$\ln \mathbf{P}_i(k) = \sum_{t=1}^{k} \frac{\ln P_i(l)}{k}. \tag{3.5}$$

In the infinite measurement case, we write

$$\ln \mathbf{P}_i = \lim_{k\to\infty} \sum_{t=1}^{k} \frac{\ln P_i(t)}{k}.$$

and the variance $\lim_{k\to\infty} \sigma^2(k) = 0$.

Proposition 3.2.4 (Expressions for posteriors) *In the case of k repeated measurements, the regional posterior for sensor i about region W_α is*

$$\mathbb{p}(\mathbf{P}_i(k)|H_\alpha)\mathbb{P}(H_\alpha) = \frac{1}{A}\int_{W_\alpha} \frac{1}{(2\pi\sigma^2(k))^{1/2}} \times \exp\Big(- \frac{\Big(\ln \mathbf{P}_i(k) - \ln(\frac{Pd_0}{d_0 + ||q_i - y||^\rho})\Big)^2}{2\sigma^2(k)}\Big) dy,$$

and the joint regional posterior for sensors $\{1, \ldots, N\}$ about region W_α is

$$\mathbb{p}(\mathbf{P}_1(k), \ldots, \mathbf{P}_N(k)|H_\alpha)\mathbb{P}(H_\alpha) = \frac{1}{A}\int_{W_\alpha} dy \prod_{l=1}^{N} \frac{1}{(2\pi\sigma^2(k))^{1/2}} \exp\Big(- \frac{\Big(\ln \mathbf{P}_l(k) - \ln(\frac{Pd_0}{d_0 + ||q_l - y||^\rho})\Big)^2}{2\sigma^2(k)}\Big).$$

Proof: Call $z = \ln \mathbf{P}_i(k)$. We compute

$$\begin{aligned} \mathbb{p}(z|H_\alpha) &= \frac{d}{dz}\frac{\mathbb{P}(Z \le z, H_\alpha)}{\mathbb{P}(H_\alpha)} = A\frac{d}{dz}\frac{\int_{-\infty}^{z}\int_{W_\alpha} \mathbb{p}(z|y)p(y)dydz}{A_\alpha} \\ &= A\frac{d}{dz}\frac{\int_{-\infty}^{z}\int_{W_\alpha} \big(\mathbb{p}(z|y)/A\big)dydz}{A_\alpha} = \frac{\int_{W_\alpha} \mathbb{p}(z|y)dy}{A_\alpha}. \end{aligned}$$

Since $z = \ln \mathbf{P}_i(k) = \sum_{t=1}^{k} \frac{\ln P_i(t)}{k}$, the conditional probability is

$$\mathbb{p}(z|y) = \frac{1}{(2\pi\sigma^2(k))^{1/2}} \exp\Big(- \frac{\Big(\ln \mathbf{P}_i(k) - \ln(\frac{Pd_0}{d_0 + ||q_i - y||^\rho})\Big)^2}{2\sigma^2(k)}\Big) dy.$$

The regional posterior is

$$\mathbb{p}(\mathbf{P}_i(k)|H_\alpha)\mathbb{P}(H_\alpha) = \frac{\int_{W_\alpha} \frac{1}{(2\pi\sigma^2(k))^{1/2}} \exp\Big(- \frac{\big(\ln \mathbf{P}_i(k) - \ln(\frac{Pd_0}{d_0 + ||q_i - y||^\rho})\big)^2}{2\sigma^2(k)} \Big) dy}{A_\alpha} \times \frac{A_\alpha}{A}$$

$$= \frac{\int_{W_\alpha} \frac{1}{(2\pi\sigma^2(k))^{1/2}} \cdot \exp\Big(- \frac{\big(\ln \mathbf{P}_i(k) - \ln(\frac{Pd_0}{d_0 + ||q_i - y||^\rho})\big)^2}{2\sigma^2(k)} \Big) dy}{A}.$$

Equations for the joint regional posterior follow by independence of measurements.

3.2.3 Geometric properties of regional source localization

In this section we derive geometric properties of the regional localization problem in the infinite measurements limit. These properties allow us to conclude the following two results. First, for certain source locations, a single sensor suffices to asymptotically detect the correct hypothesis. Second, for the asymptotic detection problem with two sensors, Voronoi partitions ensure zero probability of error. These results should be viewed against the fact that for exact localization, even in the noise-free case, at least three non-collinear sensors are needed for detection.

Remark 3.2.5 (Asymptotic density) *For infinite measurement, the conditional* *probability density is*

$$\begin{aligned}\mathrm{p}(\mathbf{P}_i|y) &= \lim_{k\to\infty}\frac{1}{(2\pi\sigma^2(k))^{1/2}}\exp\Big(-\frac{\Big(\ln\mathbf{P}_i-\ln(\frac{Pd_0}{d_0+\|q_i-y\|^\rho})\Big)^2}{2\sigma^2(k)}\Big)\\ &=\delta_{\mathrm{D}}\Big(\ln\mathbf{P}_i-\ln\frac{Pd_0}{d_0+\|q_i-y\|^\rho}\Big).\end{aligned}$$ □

Before presenting the next result, we introduce two shorthands. First, given a region W and a disk B, we call arclength$(W\cap\partial B)$ the sum of the length of the arcs in the set $W\cap\partial B$. Second, let $B(q,r)$ denote the disk centered at q with radius r.

Lemma 3.2.6 (The arc length property) *The asymptotic regional posterior for sensor i about region W_α is*

$$\mathrm{p}(\mathbf{P}_i|H_\alpha)\mathbb{P}(H_\alpha)=\frac{d_0+r_i^\rho}{A\rho r_i^{\rho-1}}\,\mathrm{arclength}\big(W_\alpha\cap\partial B(q_i,r_i)\big),$$

where $r_i=(\frac{P}{P_i}-1)^\rho$ and, as usual, q_i is the position of sensor i, d_0 is the nominal distance, ρ is the path loss exponent, and A is the environment area.

The proof of the lemma is given in the Appendix A. This lemma can be interpreted as follows. The asymptotic average measurement $\mathbf{P}_i$ determines the circle of radius r_i centered in q_i where the source lies. The posterior $\mathrm{p}(\mathbf{P}_i|H_\alpha)\mathbb{P}(H_\alpha)$ is used in MAP algorithms and is proportional to the arclength.

Next, we consider the case in which $M=N=2$, that is, there are only two sensors and two regions of interest. In the following lemma, we show that Voronoi

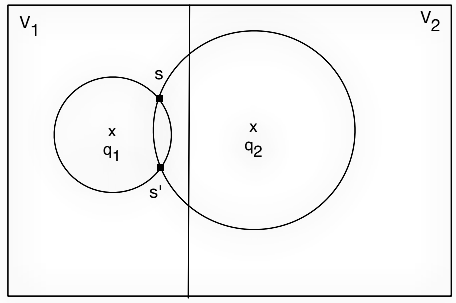


Figure 3.1: This figure shows two nodes q_1 and q_2, with a source $s \in V_1$.

partitions ensure correct detection almost surely for two sensors implementing the MAP estimation algorithm.

Lemma 3.2.7 (Optimality of Voronoi partitions) *In an environment Q with two sensors at distinct locations q_1 and $q_2 \in Q$, consider the regional localization problem with Voronoi partition (V_1, V_2) generated by q_1 and q_2. In the limit of infinite measurements, the MAP localization algorithm finds the region containing $s \in Q$ almost surely.*

Proof: The result is a consequence of Lemma 3.2.6. The intersection of $\big(\partial B(q_1, r_1) \cap \partial B(q_2, r_2)\big)$ where $r_i = (\frac{P}{P_i} - 1)^\rho$ for $i \in \{1, 2\}$, is a finite set with 2 elements almost surely, i.e., as long as the source does not belong to the line through q_1 and q_2. Accordingly, define s and s' to be these two points. The joint conditional probability, $\mathbb{p}(\mathbf{P}_1, \mathbf{P}_2|y) = \mathbb{p}(\mathbf{P}_1|y)\mathbb{p}(\mathbf{P}_2|y) = \delta_{\mathrm{D}}\Big(\ln \mathbf{P}_1 - \ln \frac{Pd_0}{d_0 + \|q_1 - y\|^\rho}\Big)\delta_{\mathrm{D}}\Big(\ln \mathbf{P}_2 -$

$\ln \frac{Pd_0}{d_0+\|q_2-y\|^\rho}\Big)$, is non-zero only at the two points s and s'. Therefore, by using the total probability theorem, the asymptotic regional joint posterior for sensors 1 and 2 about region V_i, $i \in \{1,2\}$, is

$$\mathbb{p}(\mathbf{P}_1, \mathbf{P}_2|H_i)\mathbb{P}(H_i) := \begin{cases} 0, & \text{if } s, s' \notin V_i, \\ \\ \mathbb{p}(\mathbf{P}_1, \mathbf{P}_2), & \text{otherwise} \end{cases}$$

Either s or s' has to be the source. Assume s is the source, we will show that if $s \in V_1$, then the joint regional posterior can be positive only in V_1. In fact, $s' \notin Q$ implies $\mathbb{p}(\mathbf{P}_1, \mathbf{P}_2|H_2)\mathbb{P}(H_2) = 0$ because $\mathbb{p}(\mathbf{P}_1, \mathbf{P}_2|y) = 0$ for all $y \in V_2$. If on the other hand $s' \in Q$, then by the definition of s and s', $\|q_1 - s\| = \|q_1 - s'\|$ and $\|q_2 - s\| = \|q_2 - s'\|$, and by the definition of Voronoi partitions (Def. 3.1.2) we know that $s \in V_1$ implies $s' \in V_1$. This completes the proof that for Voronoi partitions in the two sensors case, $\mathbb{p}(\mathbf{P}_1, \mathbf{P}_2|H_i)\mathbb{P}(H_i)$ is non-zero only in the correct region. The two lemmas presented in this section have interesting implications. Lemma 3.2.6 implies that, for certain source locations and as the noise becomes smaller, the MAP estimation algorithm can determine the correct region containing the source with a unique sensor. That is true when the circle centered at a sensor location with radius r_i is contained in the region W_i. Lemma 3.2.7 on the other hand, gives one example where the selection of Voronoi partitions makes it possible to locate the source with only two sensors.

3.2.4 Asymptotic properties of regional source localization

We show here some properties of the MAP algorithm when applied to regional source localization for a general number of sensors and regions. We start by presenting a property of non-collinear sensors when applied to source localization using measurements undergoing log-normal shadowing.

Lemma 3.2.8 (Three non-collinear sensors) *For $d_0 > 0$ and $\rho > 0$, given a source $s \in \mathbb{R}^2$ and three non-collinear sensors q_1, q_2 and $q_3 \in \mathbb{R}^2$, the only solution for the equation $\sum_{i=1}^{3} \left(\ln \frac{d_0+\|z-q_i\|^\rho}{d_0+\|s-q_i\|^\rho} \right)^2 = 0$ is $z = s$.*

Proof: In fact, it is easy to check that the sum is zero at $z = s$. Uniqueness of this solution is verified by noting that the sum of the square terms is zero only if all the summands are zero. Let $q = (x, y)$ and $q_i = (q_{i1}, q_{i2})$. The solution $z = s$ is unique if and only if the following system has a unique solution:

$$\begin{bmatrix} -2(q_{11} - q_{21}) & -2(q_{12} - q_{22}) \\ -2(q_{11} - q_{31}) & -2(q_{12} - q_{32}) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}, \tag{3.6}$$

where k_1 and k_2 are known values determined by the measurements and the positions of the sensors. The system presented in Equation (3.6) has a unique solution if and only if the system is consistent and the determinant of the matrix is non zero, i.e., the three points are non-collinear. As usual, assume that N sensors are at positions q_i, $i \in \{1, \ldots, N\}$ and that the environment is partitioned into closed regions. For a region W_α, define the two scalar quantities

$$U_\alpha = \max_{\substack{y\in W_\alpha \\ i\in\{1,\ldots,N\}}} \left| \ln \frac{d_0 + \|y - q_i\|^\rho}{d_0 + \|s - q_i\|^\rho} \right|, \tag{3.7}$$

$$L_\alpha = \min_{y\in W_\alpha} \sum_{i=1}^{N} \left(\ln \frac{d_0 + \|y - q_i\|^\rho}{d_0 + \|s - q_i\|^\rho} \right)^2. \tag{3.8}$$

Both quantities are well posed because they are the maximum and minimum value of a continuous function over a compact domain. Additionally, U_α is strictly positive for all source locations $s \in Q$ and L_α is strictly positive for all source locations $s \in Q \setminus W_\alpha$. The latter statement follows from Lemma 3.2.8 and from the fact that the distance from s to W_α is strictly positive for all $s \notin W_\alpha$. Define

$$\eta_\alpha = \sqrt{U_\alpha^2 + \frac{L_\alpha}{2N}} - U_\alpha > 0, \tag{3.9}$$

for all $s \notin W_\alpha$. We state the following result on the magnitude of sums of powers.

Lemma 3.2.9 (On the posterior of a wrong hypothesis) *Consider L_α, U_α and η_α as defined in* (3.7), (3.8) *and* (3.9). *Assume the source s is outside W_α and the noise n_i satisfies $|n_i| \le \eta_\alpha$ for all $i \in \{1,\ldots,N\}$ and $\alpha \in \{1,\ldots,M\}$. The following statements hold:*

1. *the joint measurement is lower bounded as*

$$\min_{y\in W_\alpha} \sum_{i=1}^{N} \left(\ln \frac{d_0 + \|y - q_i\|^\rho}{d_0 + \|s - q_i\|^\rho} + n_i \right)^2 \ge \frac{1}{2} L_\alpha, \text{ and}$$

2. the posterior probability for the wrong hypothesis α is upper bounded as

$$\mathbb{p}(P_1, \ldots, P_N|H_\alpha)\mathbb{P}(H_\alpha) \le \frac{A_\alpha \exp\bigl(-L_\alpha/4\sigma^2\bigr)}{A(2\pi\sigma^2)^{N/2}}.$$

Proof: To prove the first statement, consider the expansion

$$\left(\ln\frac{d_0+\|y-q_i\|^\rho}{d_0+\|s-q_i\|^\rho}+n_i\right)^2 = \left(\ln\frac{d_0+\|y-q_i\|^\rho}{d_0+\|s-q_i\|^\rho}\right)^2+2\left(\ln\frac{d_0+\|y-q_i\|^\rho}{d_0+\|s-q_i\|^\rho}\right)n_i+\sum_{i=1}^N n_i^2.$$

By computing lower bounds for each term and substituting the definition of η_α, we obtain

$$\begin{aligned}
&\min_{y\in W_\alpha}\sum_{i=1}^N\left(\ln\frac{d_0+\|y-q_i\|^\rho}{d_0+\|s-q_i\|^\rho}+n_i\right)^2 \ge L_\alpha - 2NU_\alpha\eta_\alpha - N\eta_\alpha^2\\
&= L_\alpha + 2NU_\alpha^2 - 2NU_\alpha\sqrt{U_\alpha^2+\frac{L_\alpha}{2N}} - N\left(U_\alpha^2+\frac{L_\alpha}{2N}\right) - NU_\alpha^2 + 2NU_\alpha\sqrt{U_\alpha^2+\frac{L_\alpha}{2N}}\\
&= \frac{1}{2}L_\alpha.
\end{aligned}$$

The second statement follows directly from the first statement because of the equality

$$\ln P_i - \ln\frac{Pd_0}{d_0+\|y-q_i\|^\rho} = \ln\frac{d_0+\|y-q_i\|^\rho}{d_0+\|s-q_i\|^\rho}+n_i,$$

and because of the fact that the surface integral of a function f is upper bounded by the surface integral of the maximum value of f. We are now ready for the convergence theorem. We introduce the standard function $Q:\mathbb{R}\to\mathbb{R}_{>0}$ by

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty}\exp(-y^2/2)dy.$$

Theorem 3.2.10 (Elimination of wrong hypothesis) *Consider sensors at positions* $q_1, \ldots, q_N$*. Let* σ *be the noise variance. If the source* $s \notin W_\alpha$*, then*

$$\mathbb{P}\Big[\mathbb{p}(P_1, \ldots, P_N | H_\alpha)\mathbb{P}(H_\alpha) \leq \epsilon_\alpha(\sigma)\Big] \geq \mu_\alpha(\sigma),$$

where

$$\epsilon_\alpha(\sigma) = \frac{A_\alpha \exp(-L_\alpha/4\sigma^2)}{A(2\pi\sigma^2)^{N/2}}, \quad \mu_\alpha(\sigma) = (1 - 2Q(\eta_\alpha/\sigma))^N.$$

Furthermore, in the k *repeated measurement case, if at least* 3 *sensors are non-collinear, then* $\lim_{k\to\infty} \epsilon_\alpha(\sigma_k) = 0^+$ *and* $\lim_{k\to\infty} \mu_\alpha(\sigma_k) = 1^-$*.*

Proof: From Lemma 3.2.9, we compute

$$\mathbb{P}\big[\mathbb{p}(P_1, \ldots, P_N | H_\alpha)\mathbb{P}(H_\alpha) \leq \epsilon_\alpha(\sigma)\big] \geq \mathbb{P}\left[[n_1, \ldots, n_N]^T \in [-\eta_\alpha, \eta_\alpha]^N\right]$$
$$= \prod_{i=1}^{N} \left(\frac{1}{2} - \mathbb{P}[n_i > \eta_\alpha] + \frac{1}{2} - \mathbb{P}[n_i < -\eta_\alpha]\right) = \big(1 - 2Q(\eta_\alpha/\sigma)\big)^N.$$

The first inequality follows from the fact that Lemma 3.2.9 holds whenever all $|n_i| \leq \eta_\alpha$. The proofs of the two limits of $\lim_{k\to\infty} \epsilon_\alpha(\sigma_k)$ and $\lim_{k\to\infty} \mu_\alpha(\sigma_k)$ are immediate when there are at least 3 non-collinear sensors. Indeed, if there are at least 3 non-collinear sensors and if $s \notin W_\alpha$, then Lemma 3.2.8 applies and one can show $L_\alpha > 0$ and $\eta_\alpha > 0$. This theorem states that, as $\sigma \to 0^+$, the joint regional posterior

$\mathbb{p}(P_1, \ldots, P_N | H_\alpha)\mathbb{P}(H_\alpha)$ takes an arbitrarily small value with a probability that goes arbitrarily close to 1 when H_α is not the correct hypothesis. This is so as

$Q(x) \to 0$ as $x \to \infty$. To complement the Theorem 3.2.10, we prove below that for the correct hypothesis, the probability density is lower bounded by a positive term with probability one.

Theorem 3.2.11 (Strict positivity of correct hypothesis) *Consider sensors at positions $q_1, \dots, q_N$. Let σ be the noise variance. If the source $s \in W_{\overline{\alpha}}$, then*

$$\mathbb{P}\left[\mathbb{p}(P_1, \dots, P_N | H_{\overline{\alpha}})\mathbb{P}(H_{\overline{\alpha}}) \geq \Psi(\sigma)\right] \geq \Omega(\sigma),$$

where

$$\Psi(\sigma) = \mathbb{p}(P_1, \dots, P_N) - \sum_{\substack{\alpha=1,\dots,M \\ \alpha \neq \overline{\alpha}}} \frac{A_\alpha \exp(-L_\alpha / 4\sigma^2)}{A(2\pi\sigma^2)^{N/2}},$$

$$\Omega(\sigma) = \prod_{\substack{\alpha=1,\dots,M \\ \alpha \neq \overline{\alpha}}} \mu_\alpha(\sigma) = \prod_{\substack{\alpha=1,\dots,M \\ \alpha \neq \overline{\alpha}}} (1 - 2Q(\eta_\alpha/\sigma))^N.$$

Furthermore, in the k repeated measurement case, if at least 3 sensors are non-collinear, then $\lim_{k\to\infty} \Psi(\sigma_k) = \mathbb{p}(P_1, \dots, P_N) > 0$ and $\lim_{k\to\infty} \Omega(\sigma_k) = 1^-$.

Proof: The proof of this theorem follows directly from Theorem 3.2.10 and from the total probability theorem. Call $z = [P_1, \dots, P_N]^T$. We know from the total probability theorem that

$$\mathbb{p}(z) = \sum_{\alpha=1}^{M} \mathbb{p}(z|H_\alpha)\mathbb{P}(H_\alpha) = \mathbb{p}(z|H_{\overline{\alpha}})\mathbb{P}(H_{\overline{\alpha}}) + \sum_{\substack{\alpha=1,\dots,M \\ \alpha \neq \overline{\alpha}}} \mathbb{p}(z|H_\alpha)\mathbb{P}(H_\alpha)$$

and, in turn, that

$$\mathbb{p}(z|H_{\overline{\alpha}})\mathbb{P}(H_{\overline{\alpha}}) = \mathbb{p}(z) - \sum_{\substack{\alpha=1,\dots,M \\ \alpha\neq\overline{\alpha}}} \mathbb{p}(z|H_\alpha)\mathbb{P}(H_\alpha).$$

From Theorem 3.2.10

$$\mathbb{P}\left[\mathbb{p}(z|H_{\overline{\alpha}})\mathbb{P}(H_{\overline{\alpha}}) \geq \mathbb{p}(z) - \sum_{\substack{\alpha=1,\dots,M \\ \alpha\neq\overline{\alpha}}} \epsilon_\alpha(\sigma)\right]$$
$$\geq \prod_{\substack{\alpha=1,\dots,M \\ \alpha\neq\overline{\alpha}}} \mathbb{P}\Big[\mathbb{p}(z|H_\alpha)\mathbb{P}(H_\alpha) \leq \epsilon_\alpha(\sigma)\Big] \geq \prod_{\substack{\alpha=1,\dots,M \\ \alpha\neq\overline{\alpha}}} \mu_\alpha(\sigma)$$

As $\lim_{k\to\infty}\sigma_k = 0^+$, $\lim_{k\to\infty}\Psi(\sigma_k) = \mathbb{p}(z)$ and $\lim_{k\to\infty}\Omega(\sigma_k) = 1^-$. Theorem 3.2.11 complements Theorem 3.2.10 in that is shows that as $\lim_{k\to\infty}\sigma_k = 0^+$, the largest regional posterior is the one associated with the correct hypothesis.

Remark 3.2.12 (Almost sure convergence of MAP) *Using a MAP algorithm to solve the problem of regional localization, is assured to provide a correct answer, almost surely, in the limit of a large number of measurements. This follows directly from Theorem 3.2.11 and Theorem 3.2.10.*

3.3 Distributed sequential regional localization

In this section we assume that each sensor is a processor that can perform computational tasks as well as communicate to other processors according to a specified communication graph. Each processor takes measurements and computes a conditional posterior that it communicates to all its neighbors and then makes a decision if a desired accuracy is reached. A group of regions is associated with each processor. The processor will need to provide a decision about which of these regions if any contains the source. We call such a group the regions of responsibility of the processor. We do not assume any constraints on the assignment of regions of responsibilities. We present the algorithm in Subsection 3.3.1 and describe its properties in Subsection 3.3.2.

3.3.1 Distributed algorithm based on sequential sensing, communication and hypothesis testing

We present below a distributed algorithm where each processor decides whether or not its region of responsibility contains the source. The algorithm, as presented, has a predefined number of measurements that need to be taken by each processor.

For each processor $i \in \{1, \ldots, N\}$, the set of neighbors $\mathcal{N}_i$ consists of the processor itself along with the processors that can communicate with it. The ith processor is responsible for a set $\mathcal{R}_i$ of M_i regions. We denote these M_i regions by W_α for $\alpha \in \mathcal{R}_i$. The processor collects the measurements from its neighboring processors, and calculates two posteriors for all regions W_α, $\alpha \in \mathcal{R}_i$. The first posterior corresponds to the hypothesis that the source is in W_α, the second posterior corresponds to the hypothesis that the source is outside W_α. Once the processor reaches a pre-defined level of confidence, it provides a decision about whether or not W_α contains the source. The ith processor stops running its test when it reaches a decision about all W_α, $\alpha \in \mathcal{R}_i$. The processor then sets its decision to either `yes`, the source is in W_α, or `no`, no source is in $\cup_{\alpha \in \mathcal{R}_i} W_\alpha$. Each processor continues to sense and transmit its measurements until all its neighbors $j \in \mathcal{N}_i$ have reached a decision. We give here a formal description of the algorithm.

Algorithm : *Sense, Transmit & Test*

algorithm tolerance: $0 < \epsilon \ll \frac{1}{2}$

network processors: $i \in \{1, \ldots, N\}$

regions: W_α, $\alpha \in \{1, \ldots, M\}$

state of processor i contains:

$\texttt{a-dcsn}_i \in \{\texttt{yes}\ \textit{source} \in W_\alpha, \texttt{no}\ \textit{source} \in \cup_\alpha W_\alpha, \texttt{unknown}\}$,

for all $j \in \mathcal{N}_i : q_j, \texttt{a-stop}_j \in \{\texttt{false}, \texttt{true}\}$,

for all $\alpha \in \mathcal{R}_i : W_\alpha, \texttt{r-stop}_\alpha \in \{\texttt{false}, \texttt{true}\}$,

$\texttt{r-dcsn}_\alpha \in \{\texttt{yes}, \texttt{no}, \texttt{unknown}\}$

Processor i with set of neighbors $\mathcal{N}_i$ executes:

1: transmit q_i to $j \in \mathcal{N}_i$

set $k := 0$, $\texttt{a-stop}_i := \texttt{false}$ and $\texttt{a-dcsn}_i := \texttt{unknown}$

set $\texttt{r-stop}_\alpha := \texttt{false}$ and $\texttt{r-dcsn}_\alpha := \texttt{unknown}$ for $\alpha \in \mathcal{R}_i$

While $\exists\, j \in \mathcal{N}_i$ with $\texttt{a-stop}_j == \texttt{false}$ **do**

  update $k := k+1$ and take measurement $P_i(k)$

  compute $\ln \mathbf{P}_i(k) = \sum_{t=1}^{k} \frac{\ln \mathbf{P}_i(t)}{k}$

  transmit $\ln \mathbf{P}_i(k)$ to $j \in \mathcal{N}_i$

  store $\ln \mathbf{P}_{\mathcal{N}_i}(k) = \{\ln \mathbf{P}_i(k)\} \cup \{\ln \mathbf{P}_j(k) \text{ for all } j \in \mathcal{N}_i\}$

  For all $\alpha \in \mathcal{R}_i$ with $\texttt{r-stop}_\alpha == \texttt{false}$ **do**

    If $\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k) | s \in W_\alpha) \mathbb{P}(s \in W_\alpha) > (1-\epsilon)\ \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k))$

      $\texttt{dcsn}_\alpha := \texttt{yes}, \texttt{r-stop}_\alpha := \texttt{true},\ \texttt{a-dcsn}_i := \texttt{true}$

    If $\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k) | s \notin W_\alpha) \mathbb{P}(\notin W_\alpha) > (1-\epsilon)\ \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k))$

      $\texttt{r-dcsn}_\alpha := \texttt{no}$ and $\texttt{r-stop}_\alpha := \texttt{true}$

  End For

If $\texttt{r-stop}_\alpha == \texttt{true}$ **for all** $\alpha \in \mathcal{R}_i$

  $\texttt{a-stop}_i := \texttt{true}$

  If $\texttt{a-dcsn}_i == \texttt{unknown}$

    $\texttt{a-dcsn}_i := \texttt{no}$

transmit $\texttt{a-stop}_i$ to all $j \in \mathcal{N}_i$

return $\texttt{a-dcsn}_i$

End While

3.3.2 Properties of *Sense, Transmit & Test*

We present below properties involving the accuracy and decision time of the *Sense, Transmit & Test* algorithm.

Theorem 3.3.1 (Accuracy and decision time for *Sense, Transmit & Test* algorithm) *Assume that only one source exists in the environment Q, that each processor has at least 2 neighboring processors with which it forms a non-collinear triplet, and that each processor is assigned M_i regions. Given an accuracy $\epsilon \in (0, \frac{1}{2})$, the* Sense, Transmit & Test *algorithm enjoys the following two properties:*

1. *the algorithm ends in a finite time, and*

2. *each processor i has a probability of error no larger than $2M_i\epsilon$ if $2 \leq M_i \leq 1 + \frac{1}{\epsilon}$, and no larger than ϵ if $M_i = 1$.*

Proof: It is well known [3, 5] that given two hypothesis H_1 and H_0 with known posteriors, $P(H_1)$ and $P(H_0)$, a hypothesis test that ensures that the decision under hypotheses H_0 and H_1 is correct with a probability greater than τ_0 and τ_1 respectively is the following:

1. Calculate $\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1), \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0)$

2. if $\frac{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1)}{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0)} \geq \frac{\tau_1}{1-\tau_1}$ decide in favor of H_1,

3. if $\frac{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1)}{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0)} \leq \frac{1-\tau_0}{\tau_0}$, decide in favor of H_1,

4. otherwise repeat measurements and go to 1.

We show below that the *Sense, Transmit & Test* algorithm satisfies the description above. Applying the total probability theorem we get

$$\begin{aligned}
&\mathbb{p}(P_{\mathcal{N}_i}|s \notin W_\alpha)\mathbb{P}(s \notin W_\alpha) = \mathbb{p}(P_{\mathcal{N}_i}|y \in Q)\mathbb{P}(s \in Q) - \mathbb{p}(P_{\mathcal{N}_i}|s \in W_\alpha)\mathbb{P}(s \in W_\alpha) \\
&= \mathbb{p}(P_{\mathcal{N}_i}) - \mathbb{p}(P_{\mathcal{N}_i}|s \in W_\alpha)\mathbb{P}(s \in W_\alpha). \qquad (3.10)
\end{aligned}$$

Call ($H_1 := s \in W_\alpha$) and ($H_0 := s \notin W_\alpha$). If we set $\tau_0 = \tau_1 = (1-\epsilon)$, and the thresholds to accept a hypothesis H_1, to be

$$\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1) \geq \tau_1 \mathbb{p}(P_{\mathcal{N}_i}),$$

and the thresholds to reject a hypothesis to be

$$\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0) \geq \tau_0 \mathbb{p}(P_{\mathcal{N}_i}),$$

then using Eq. (3.10), one can show that

$$\begin{aligned}
&\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1) \geq \tau_1 \ \mathbb{p}(P_{\mathcal{N}_i}) \ \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0) \leq (1-\tau_1)\mathbb{p}(P_{\mathcal{N}_i}) \\
&\Rightarrow \frac{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1)}{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0)} \geq \frac{\tau_1}{1-\tau_1}. \qquad (3.11)
\end{aligned}$$

Similarly, assuming H_0 is correct, one can show that

$$\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0) \geq \tau_0 \ \mathbb{p}(P_{\mathcal{N}_i}) \Rightarrow \frac{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1)}{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0)} \leq \frac{1-\tau_0}{\tau_0}. \qquad (3.12)$$

Assuming H_1 is correct, the probability of correct decision for the ith processor is no smaller than τ_1, for each of the regions $W_\alpha, \alpha \in \mathcal{N}_i$. Similar result hold assuming H_0 is correct.

The maximum number of errors that a processor can make in a decision is two: a mis-detection and a false-alarm for W_α where $\alpha \in \mathcal{R}_i$. Alternatively all the other combinations of choices result in at most one error, since the ith processor can declare at most one hypothesis H_α to be correct for all $\alpha \in \mathcal{R}_i$.

The scenarios where the decision of the processor is erroneous are presented below:

1. If one of the regions W_α satisfies $\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1) \geq \tau_1 \mathbb{p}(P_{\mathcal{N}_i})$, then for all $\beta \in \mathcal{R}_i \setminus \alpha$, the following holds (from the complete probability theorem)

$$\begin{aligned}&\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|s \in W_\beta)\mathbb{P}(s \in W_\beta) < (1-\tau_1)\mathbb{p}(P_{\mathcal{N}_i})\\ \Rightarrow &\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|s \notin W_\beta)\mathbb{P}(s \notin W_\beta) \geq \tau_1 \ \mathbb{p}(P_{\mathcal{N}_i}) = \tau_0 \mathbb{p}(P_{\mathcal{N}_i}).\end{aligned} \tag{3.13}$$

   From Eqs. (3.12) and (3.13) it follows that the source can be detected in at most one region W_α. It follows that at most one false alarm can happen, which might or might not be accompanied with one mis-detection.

2. If none of the regions of responsibilities of the processor contain the source, then the processor can make at most one mistake by having at most one false alarm.

To write a formal proof, we introduce p_f and p_m to be the probability of false alarm and mis-detection, where p_f corresponds to choosing `yes` while the correct decision is `no` and mis-detection corresponds to choose `no` when the correct decision is `yes` for any region W_α. Here that $p_f = p_m = \epsilon$. A processor makes an error if it wrongly decides `yes`/`no` on W_α for any $\alpha \in \mathcal{R}_i$. Following the analysis above, the probability of error for the ith processor is:

$$P_e < \binom{M_i}{1}\left(p_m P(s \notin \cup_{\alpha\in\mathcal{R}_i} W_\alpha) + (p_f + \binom{M_i-1}{1} p_f p_m) P(s \in \cup_{\alpha\in\mathcal{R}_i} W_\alpha)\right) \leq 2M_i\epsilon,$$

if $\epsilon(M_i-1) \leq 1$. If the processor has only one region of responsibility, it is straightforward to see that the processor has a probability of error no larger than ϵ.

We show now that the test ends after a finite number of measurements. For a region W_α, processor needs to decide whether the source is in W_α (H_1) or outside it (H_0).

Without loss of generality, assume that H_1 is correct for a region W_α. We know from Theorem 3.2.10 that

$$\lim_{k\to\infty} \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0) = 0^+,$$

almost surely. We also know from Theorem 3.2.11 that

$$\lim_{k\to\infty} \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1) = \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)) > 0,$$

almost surely. This has the following implication

$$\lim_{k\to\infty} \frac{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1)}{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k))} = 1,$$

which implies that for all $\bar{\epsilon} > 0$, there exists $0 < K < \infty$, s.t.

$$\begin{aligned}
&|\frac{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K)|H_1)\mathbb{P}(H_1) - \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K))}{\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K))}| < \bar{\epsilon} \\
&\iff -\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K)|H_1)\mathbb{P}(H_1) + \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K)) < \bar{\epsilon}\ \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K)) \\
&\iff \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K)|H_1)\mathbb{P}(H_1) > (1-\bar{\epsilon})\ \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(K)).
\end{aligned}$$

So for any $\frac{1}{2} < \tau < 1$, there exists, almost surely, $K < \infty$, s.t.

$$\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_1)\mathbb{P}(H_1) > \tau \; \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)),$$

here $\tau = 1 - \bar{\epsilon}$, where $0 < \bar{\epsilon} < \frac{1}{2}$.

Similarly one can prove that if H_0 is correct, then there exists, almost surely, $K < \infty$, such that

$$\mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)|H_0)\mathbb{P}(H_0) > \tau \; \mathbb{p}(\mathbf{P}_{\mathcal{N}_i}(k)).$$

To complete the proof, we cover the cases where the algorithm makes a wrong decision. This is possible if the thresholds corresponding to a wrong decision are crossed at a time $K_1 < K < \infty$.

This completes the proof that the *Sense, Transmit & Test* algorithm has a finite decision time.

3.4 Numerical results

We present in this section three sets of simulations. The first two sets illustrate some properties of the *Sense, Transmit & Test* algorithm, while the third presents a modification of the algorithm that introduces an interesting extension of the work. In the first simulations, there are as many regions as there are sensors, i.e., $N = M = 10$. We start by presenting in Figure 3.2 a sample of the results obtained by the *Sense, Transmit & Test* algorithm. The figure shows the positions of the processors (equipped with sensors) as well as the partition of Q. As the partition, we adopt the Voronoi partition generated by the processors positions; each processor is responsible for its corresponding Voronoi region. As stated in the caption, after 113 observations all decisions have been made and the source has been correctly localized.

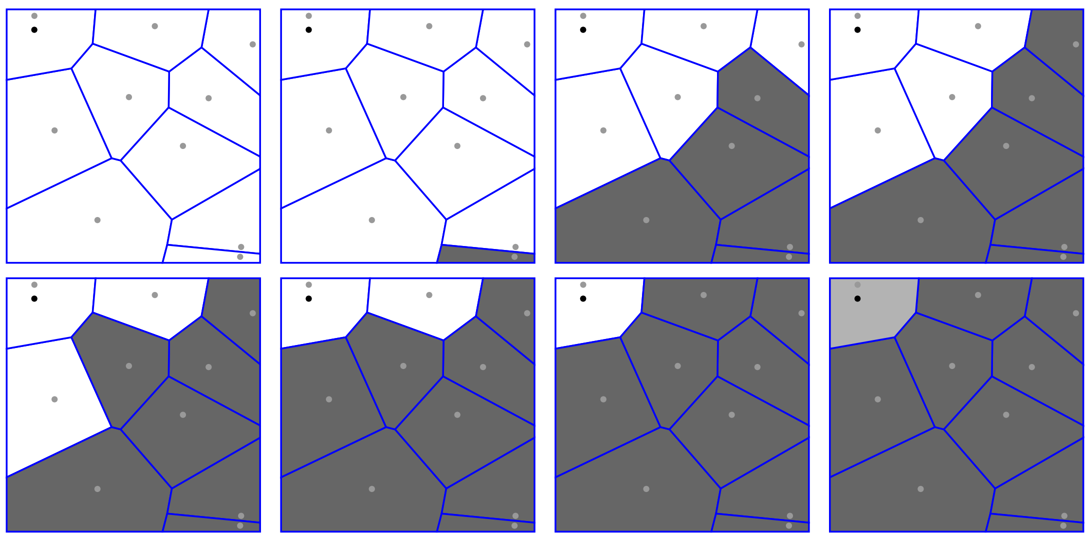


Figure 3.2: This picture illustrates an evolution of the output of the *Sense, Transmit & Test* algorithm. At each instant a region is colored in white, light gray or dark gray, indicating `unknown`, `yes` or `no` respectively. The output of the distributed algorithm is shown at times $0, 1, 3, 4, 6, 8, 11, 113$ respectively. In this run we $\epsilon = 0.01$ and $\sigma = 0.5$ with $N = M = 10$.

We then present in Figure 3.3 a plot that shows how the expected number of observations needed to reach a decision varies with the accuracy ϵ in the algorithm. Clearly, the probability of correct detection increases for decreasing values of ϵ.

In Figure 3.4 we show how the expected number of observations needed to reach a

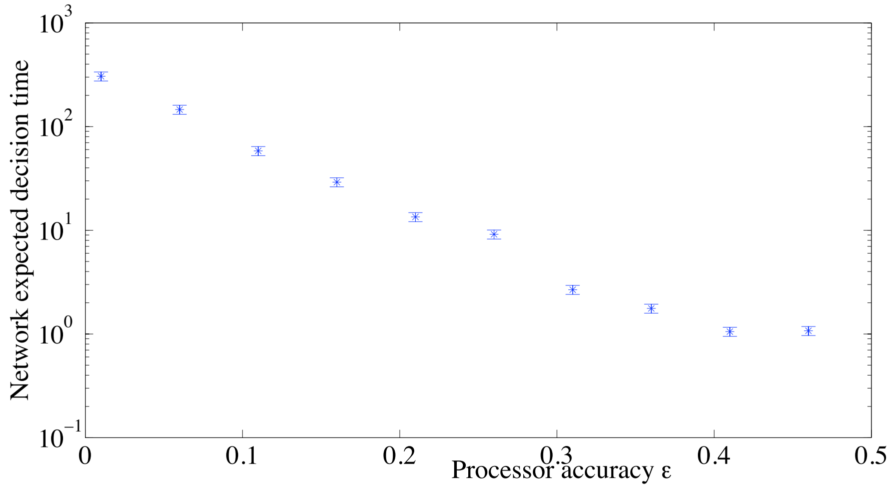


Figure 3.3: This plot shows the expected time it takes a network of 10 processors implementing the *Sense, Transmit & Test* algorithm to reach a decision for a noise standard of deviation $\sigma = 0.1$ when the probability of error ϵ varies. We show the logarithm of the decision time. Note that the network decision time seems to grow exponentially with the desired accuracy as is standard in sequential hypothesis testing. The network is assumed to have reached a decision when all processors have decided. The expected decision time is calculated over 1000 runs.

decision increases with the standard deviation of the noise.

Next, we report the second sets of simulations, where we have differing numbers of regions and sensors. Specifically, we have $N = 4$ sensors and $M = 16$ regions. Figure 3.5 illustrates the evolution of the *Sense, Transmit & Test* algorithm in this case. The overall accuracy for each processor is 0.9. This is achieved by setting $\epsilon = 0.1/8$.

In this third set of simulations, we show the output of a modified, multi-resolution

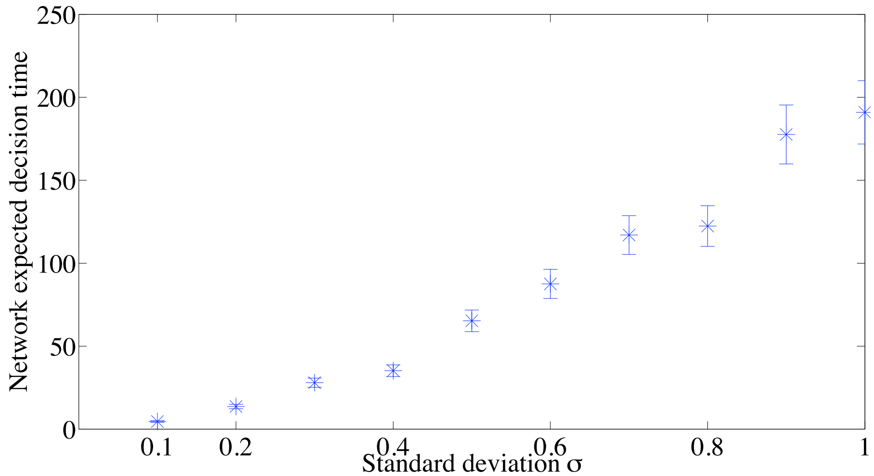


Figure 3.4: This plot shows the expected time it takes a network of 10 processors implementing the *Sense, Transmit & Test* algorithm to reach a decision with a probability of error no larger than $\epsilon = 0.01$ as the noise standard of deviation σ varies. The network is assumed to have reached a decision when all processors have decided. The expected decision time is calculated over 100 runs.

version of the *Sense, Transmit & Test* algorithm. This multi-resolution version runs over multiple stages, at each stage the environment under consideration is divided in two regions. Observations are taken at each stage until one of the two regions is rejected with an accuracy of $1 - \epsilon$. The rejected region is removed from the environment, and the remaining region is again divided in two regions. Observations are transferred from one stage to another and re-used to reach a decision about the more fine environment division. A sample output of the modified algorithm is shown in Figure 3.6. In order to reach the same precision in localization as that shown in Figure 3.5, we divide the

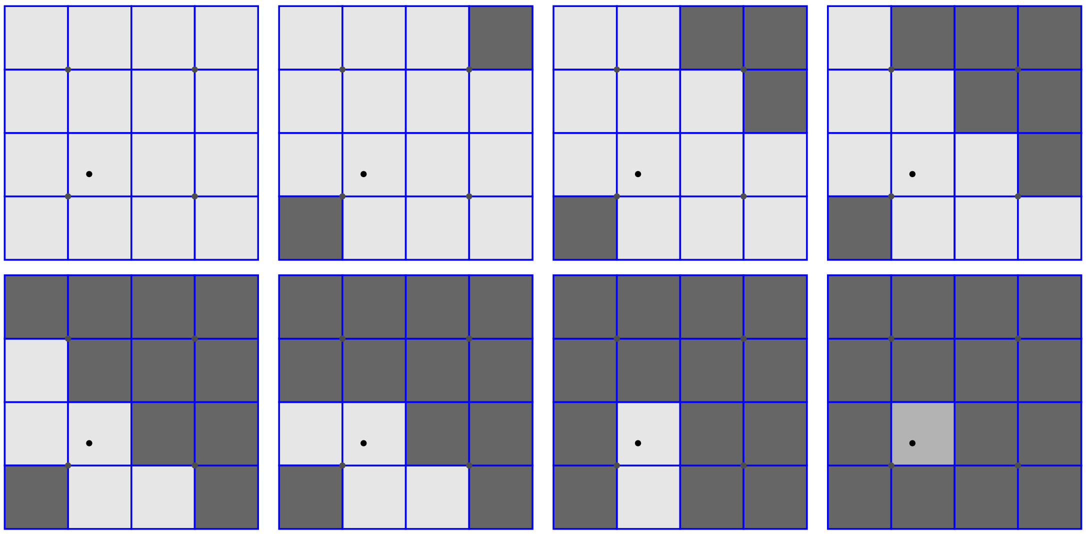


Figure 3.5: This picture illustrates an evolution of the *Sense, Transmit & Test* algorithm. At each instant a region is colored in white, light gray or dark gray, indicating `unknown`, `yes` or `no` respectively. The output of the distributed algorithm is shown at times $1, 4, 6, 7, 13, 131, 142, 202$ respectively. In this run we set $\epsilon = 0.1/8$ and $\sigma = 0.5$ and $N = 4$ and $M = 16$.

regions 4 times. Note that the original *Sense, Transmit & Test* algorithm reached its decision after an average of 290 observations whereas the multi-resolution algorithm did so after an average of 100 observations. We calculated these values from 1000 Monte-Carlo runs, that is with an error of $\pm 3\%$ to show a similar probability of error with the same level of fineness. We leave a rigorous analysis of the multi-resolution algorithm to future work.

We conclude with a general remark. The *Sense, Transmit & Test* algorithm pre-

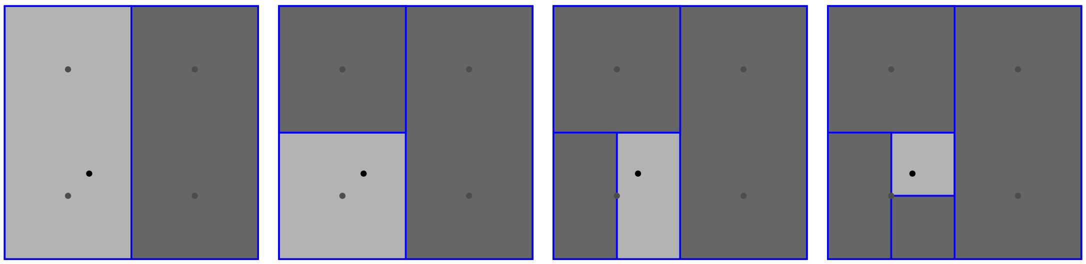


Figure 3.6: This picture illustrates an evolution of the modified version of the *Sense, Transmit & Test* algorithm. The output is shown at times $7, 11, 71, 72$ respectively. The rejected regions are shown in dark grey and the ones accepted at each stage are shown in light grey. In this run we set $\epsilon = 0.1$ and $\sigma = 0.5$ and $N = 4$ and $M = 2$ at each set of tests.

sented in this work might at first glance look similar to sequential multiple hypothesis testing algorithms by elimination such as the one presented in [49]. A closer comparison of the two algorithms shows that while in this work at most $2M$ tests are run at each sample, the hypothesis test by sequential elimination requires a number of tests of the order 2^M as it proceeds by a pairwise comparison over all hypothesis. Nonetheless, it is worth mentioning that while the sequential elimination algorithm leads to a decision as soon as all but one hypothesis is eliminated, we wait here until the last hypothesis reaches the required certainty level. This can be seen in Figure 3.2 where all but one hypothesis were canceled at the 11th observation, yet the algorithm did not end until the 113th observation when the last processor reached its required accuracy. The geometric aspects and the properties associated with the regional localization problem made it possible to propose the simpler, yet less general, *Sense, Transmit & Test* algorithm.

3.5 Conclusion

In this work, we looked at the problem of source localization in a multiple hypothesis testing setting. We based our formulation on the geometric properties of the MAP algorithm when applied to regional localization. We proved that when measurements are available from three or more non-collinear sensors, MAP based algorithms choose the correct region almost surely in the limit of a large number of measurements. We then presented a sequential distributed algorithm where each processor senses, transmits and tests to provide a decision. We analyzed the algorithm and provided a measure of its accuracy and showed that it ends in a finite time. We conclude this chapter by numerically illustrating the algorithm's performance.

There are two direct extensions for this work that we are considering. The first is using an adaptive hierarchical methods based on quadtrees [50] to increase the level of details in the choice of regions. The regions could be finely divided as fewer candidate regions are left; an example of such adaptation is shown at the end of the chapter. It would be interesting to study the trade off between the accuracy and the decision time as a function of the fine-gridding of the regions.

The second is allowing the algorithm to stop as soon as a processor decides that its region contains the source. As presented in this chapter, the algorithm has a proven accuracy performance based on the assumption that all processors reach their decisions independently of each other, and although we assume only one source, a processor will continue applying the *Sense, Transmit & Test* algorithm until it decides that its region does not contain the source even if the source was detected by one of the other processors. It will be interesting to see what happens to the accuracy if an individual can broadcast a `yes` to everyone in the group, allowing them to stop. Alternatively, as we showed in Figure 3.2, it is possible that only one hypothesis is left by elimination. It will also be interesting to analyze, if possible, the accuracy of an algorithm that makes use of such scenarios when they occur.

Chapter 4

Gossip algorithm for a class of environment partitioning problems with separable rewards and equitability constraints

4.1 Introduction

In this chapter we study a family of optimization problems that we call *separable optimization problems* and show how distributed gossip based algorithms can be used to solve this family of problems.

Finding algorithms that allow a network of cooperative agents to perform a given task has also been a topic of great value. Many problems might arise in such a setting: The agents might not have the ability to fuse all the information in the network at one place, the topology of the network itself might vary as some of the nodes might fail suddenly, the topology might not be available to all nodes, and in many times, nodes have limited computing ability. To deal with all these problems, distributed algorithms are sought.

4.1.1 Chapter contribution

The major contribution of this chapter is that it introduces and implements a distributed peer-to-peer algorithm that solves regional optimization problems where the objective is to find partitions that optimize a cost function while maintaining an equitability constraint. The cost function has the property of depending (or being transformable into a function that depends) separately on each region. The applications of this family of problems are numerous. We study in this chapter four different problems in details. We start with equitable partitioning problems and present a distributed gossip algorithm that solves this problem. We then study three problems: we start with the doubly equitable partitioning problem where the objective is to find partitions that satisfy the equitability constraints under two different measures, we then study the isoperimetric problem, where for a given area the sum of the isoperimetric ratio (perimeters) of the partitions needs to be maximized (minimized) and finally we study an environment partitioning problem where the objective is to improve the performance of a source localization algorithm. We present and implement an algorithm that solves all these problems. For cases where some information might be missing, we introduce an algorithm that we implement as a solution for the doubly equitable partitioning problem. Our presented algorithms are general and can be used to solve any problem belonging to a specific family of problems that we present.

4.1.2 Notations and definitions

We present below a set of notations that will be used throughout the chapter.

1. $\lambda(W)$ denotes the measure of a region W under a given density function $\lambda(x)$.

$$\lambda(W) = \iint_W \lambda(x)\, dx,$$

   where $x \in \mathbb{R}^d$.

2. $|W|$ denotes the area of the region W.

3. $\mathrm{perim}(W)$ denotes the length of the boundary ∂W of the region W, i.e., the perimeter of W.

4. $\mathcal{P} = \{W : W \subseteq C\}$ denotes the set whose elements are all possible polygons included in C.

Definition 4.1.1 (Separability property) *A function $f(x_1, \ldots, x_N) : \mathbb{R}^N \to \mathbb{R}$ is said to be separable, if it can be written as the sum of functions $g_i(x_i) : \mathbb{R} \to \mathbb{R}$, for all*

$i \in \{1, 2, \dots, N\}$. *That is*

$$f(x_1, x_2, \dots, x_N) = \sum_{i=1}^{N} g_i(x_i).$$

4.2 Problems of interest

We formally present in this section the four problems of interest. The first problem we look at is motivated by load distribution problems. If we assume that the amount of energy required to serve a region where a group of people are randomly distributed is a function of the number of people in the region, and if we assume that agents serving these regions have identical abilities, we would like to assign equal loads to each agent. Such equal loading is mathematically equivalent to having regions of equal measures with respect to a given distribution, which in this case is the distribution of people or of objects that we intend to serve. These regions are called equitable in the literature and have been the subject of interest for many scientists from different backgrounds [37, 39, 40]. Finding partitions that satisfy this property has attracted the attention of researchers in many disciplines, for example see [40, 41] and references therein.

We start by giving a formal definition of equitable partitions.

Remark 4.2.1 (Definition of equitable partitions) *Given an environment C, a partition $\{W_1, \dots, W_N\}$ of C is said to be equitable under a measure $\lambda : C \to \mathbb{R}_{\geq 0}$, if it satisfies the following properties:*

1. $\cup_{i=1}^{N} W_i = C$.

2. $\lambda(W_i \cap W_j) = 0, \forall\ i, j \in \{1, \ldots, N\}$.

3. $\lambda(W_i) = \lambda(W_j), \forall\ i, j \in \{1, \ldots, N\}$.

Under the assumption of uniform load distribution, the measure $\lambda(W)$ corresponds to the Lebesgue measure of W, and equitable regions in this case are regions of equal areas.

□

We present below the equitable partitioning problem.

Definition 4.2.2 (Equitable partitioning problem) *Given N partitions $\{W_1 \ldots, W_N\}$, the equitable partitioning problem consists of finding a set of partitions $\{W_1', \ldots, W_N'\}$ such that the following holds:*

- $\cup_{i=1}^{N} W_i = \cup_{i=1}^{N} W_i'$,
- $\lambda(W_i' \cap W_j') = 0$ *for all* $i, j \in \{1, \ldots, N\}$,
- $\lambda(W_i) = \lambda(W_j)$ *for all* $i, j \in \{1, \ldots, N\}$.

Some applications require equitability of partitions with respect to more than one measure. We present below the problem of doubly-equitable partitions.

Definition 4.2.3 (Doubly equitable partitioning problem) *Given N partitions $\{W_1 \ldots, W_N\}$, the doubly equitable partitioning problem consists of finding a set of partitions $\{W_1', \ldots, W_N'\}$ such that the following holds:*

- $\cup_{i=1}^{N} W_i = \cup_{i=1}^{N} W'_i$,
- $\lambda(W'_i \cap W'_j) = 0$ *for all* $i, j \in \{1, \ldots, N\}$,
- $\gamma(W'_i \cap W'_j) = 0$ *for all* $i, j \in \{1, \ldots, N\}$,
- $\lambda(W_i) = \lambda(W_j)$ *for all* $i, j \in \{1, \ldots, N\}$,
- $\gamma(W_i) = \gamma(W_j)$ *for all* $i, j \in \{1, \ldots, N\}$.

In addition to requiring equal load distributions many application require minimizing the perimeters of regions. For example, in districting, ensuring privacy and security of personal properties could in some cases require building fences around the regions. In nature, honeycombs require a lot of energy from the bees to be built. This energy is better spent on filling the combs with honey rather than producing the wax necessary to build storage room for the honey. Seeking efficiency, bees solve the problem of finding partitions that cover an area while minimizing the sums of the perimeters of the partitions. It has been shown that the circle is the geometric shape that minimizes the perimeter while covering a given area. Having multiple regions in the partition and requiring avoidance of empty spaces between the partitions is a restriction that changes the solution of the problem. We present below the isoperimetric problem which is the second problem that we study in this chapter.

Definition 4.2.4 (Equitable isoperimetric problem) *Given* C*, choose* *a partition* $\{W_1, \ldots, W_N\}$ *of* C *that maximizes*

$$\sum_{j=1}^{N} \frac{4\pi |W_j|}{\mathcal{H}(W_j)^2},$$

subject to $\lambda(W_i) = \lambda(W_j)$ *for* $i, j \in \{1, \ldots, N\}$.

In 2001 Tales finally proved this conjecture in [43]. The existence of a hexagonal tiling depends on the shape and size of the region C. It is easy to see that an optimal solution for the equitable isoperimetric problem, is one where the hexagonal regions have equal areas and perimeters.

We introduce below a problem inspired by the isoperimetric problem introduced earlier.

Definition 4.2.5 (Minimal perimeter problem) *Given an environment* C*, choose a partition* $\{W_1, \ldots, W_N\}$ *of* C *that minimizes*

$$\sum_{j=1}^{N} \mathrm{perim}(W_j),$$

subject to $\lambda(W_i) = \lambda(W_j)$ *for* $i, j \in \{1, \ldots, N\}$.

The last problem we consider is inspired from regional localization. In this problem we are looking to choose the partitions $\{W_1, \ldots, W_N\}$ of an environment C so that the performance of a localization algorithm is optimized. The algorithm whose performance we are optimizing partitions the environment C into N regions and chooses one of the regions which is more likely to contain a source $s \in C$ that we are trying to localize. We introduce below a heuristic measure of the performance of the algorithm [51].

Definition 4.2.6 (Pointwise performance) *Given a compact region $C \subset \mathbb{R}^2$ containing a set of sensors positions $\{x_1, \ldots, x_N\}$, an unknown position of a source in C is modeled as a random variable s uniformly taking values in C, $d_0 > 0$ and $\beta \geq 2$, define the smooth* pointwise performance $\ell : C \to \mathbb{R}_{\geq 0}$ *by*

$$\ell(y) := E_s \left[\sum_{i=1}^{N} \left(\ln \frac{d_0 + \|x_i - y\|^\beta}{d_0 + \|x_i - s\|^\beta} \right)^2 \right]. \tag{4.1}$$

The pointwise performance $\ell(y)$ can be thought of as the measure of how far a point y seems to be from the source s when a given localization algorithm is applied. The higher the value of $\ell(y)$ when $s \neq y$, the better the performance of the algorithm at y. A measure for the worst case performance of the algorithm on C, can be thought of as the joint contribution of the worst case pointwise performances of the algorithm on each of the regions W_i. In addition to optimizing the performance of the algorithm, we require that the answer provided by the algorithm, i.e., the chosen region, be informative in the sense of limiting uncertainty in both x and y directions. We do so by associating a cost to the diameter of the region W, $\mathrm{diam}(W)$.

Definition 4.2.7 (Region diameter) *The diameter of a region W is*

$$diam W = \max_{p,q \in W} \|p - q\|_\infty.$$

For a polygon P, clearly

$$diam P = \max_{v,w \in Vertices(P)} \|v - w\|_\infty,$$

Definition 4.2.8 (Square coverage problem) *Given an environment C, choose a partition $\{W_1, \ldots, W_N\}$ of C that minimizes*

$$\sum_{j=1}^{N} diam(W_j), \quad \text{subject to } \lambda(W_i) = \lambda(W_j),$$

for $i, j \in \{1, \ldots, N\}$.

We are now ready to present the informative partitioning problem for optimal regional localization.

Definition 4.2.9 (Informative partitioning problem) *Given an environment C and a pointwise performance function $\ell : C \to \mathbb{R}_{\geq 0}$ as in Definition 4.2.6, and a number $\xi \in [0, 1]$, choose a partition $\{W_1, \ldots, W_N\}$ of C that maximizes*

$$\sum_{j=1}^{N} \left(\xi \min_{y \in W_j} \ell(y) - (1 - \xi) diam(W_j) \right),$$

subject to $\lambda(W_i) = \lambda(W_j)$ for $i, j \in \{1, \ldots, N\}$.

In these problems, the optimization variable is the environment partition. The space of partitions is not a Euclidean space. For example, partitions with polygonal regions may have an arbitrary number of polygonal vertices. It is possible to introduce a topology and a metric on the space of partitions; see [52].

4.3 Distributed gossip algorithms for environment partitioning

In this section we present algorithms that solve the various environment partitioning problems that we presented so far. We search for optimal partitioning over polygons. The algorithm updates the regions in a pairwise manner. At each communication round, a pair of regions W_i and W_j is chosen, their union $W = W_i \cup W_j$ is calculated and a new pair of regions W_i^+ and W_j^+ is chosen so that the objective function is optimized. Since we are interested in searching over partitions obtained by polygons, it is justifiable to limit the search for partitions to ones obtained by maps that split regions by intersecting them with half-planes. This allows us to change the optimization problem from one where the variables are the partitions, to one where the optimization variable is a real number θ, and where a real number c is used to meet the constraints.

We introduce the *gossip* algorithm presented in [36] as a distributed algorithm that solves the consensus problem. *Gossip* algorithms play a crucial role in the implementation of our proposed solution. We also show how *gossip* algorithms can be used to exploit properties of some optimization problems. We present below the consensus problem, followed by the gossip algorithm.

Definition 4.3.1 (Consensus problem) *Consider a connected graph $G = (V, E)$, where the vertex set V contains N nodes and E is the edge set. Let the $N \times 1$ vector $A(0)$ denote the initial state vector, where the ith entry $A_i(0)$ corresponds to the* *initial states associated with the ith node. The consensus problem consists of finding* $A_{ave} = \sum_{i=1}^{N} A_i(0)/N$.

One distributed algorithm that solves the problem in Definition 4.3.1 is given below.

Algorithm 4.3.2 (Gossip algorithm for consensus problem) *Denote by* $\mathcal{A}(g)$ *the algorithm described as follows. Denote by* g_{ij} *the probability that node* i *contacts its neighbor* j*. If this happens at a time* k*, the two nodes set their current values to the average of their states.* $A_i(k+1) = A_j(k+1) = \frac{A_i(k)+A_j(k)}{2}$*. Let* $A(k)$ *denote the value of the vector of the states at the end of time* k*, then* $A(k) = M(k)A(k-1)$*, where with probability* $\frac{1}{N}g_{ij}$ *the random matrix* $M(k)$ *has in its* ij*th entry* $M_{ij} = I - \frac{(e_i-e_j)(e_i-e_j)^T}{2}$*, where* $e_i = [0 \ldots 0\ 1\ 0 \ldots 0]^T$ *is an* $N \times 1$ *unit vector with the* i*th component equal to 1.*

Consensus problem: Distributed gossip algorithm

Network: *nodes* $\{1, \ldots, N\}$ *with connected communication graph* $G = (V, E)$

State of sensor i *is* x_i

For a predefined number of loops

1: *choose randomly* $i, j \in \{1, \ldots, N\}$ *s.t.* $(i, j) \in E$

2: *calculate* $\{x_i', x_j'\} = \{(x_i + x_j)/2, (x_i + x_j)/2\}$

3: ***Return:*** $x_i^+ = x_i'$ *and* $x_j^+ = x_j'$; $\{x_1, \ldots, x_N\} = \{x_1, \ldots, x_N\}^+$

The algorithm proposed in Algorithm 4.3.2 as a solution to the problem in Definition 4.3.1 has the following properties [36].

Lemma 4.3.3 (Properties of gossip algorithms) *Define the matrix g with components g_{ij} to be the adjacency matrix associated with the graph $G = (V, E)$ with N nodes, the matrix D to be the diagonal matrix with entries $D_i = \sum_{j=1}^{N} [g_{ij} + g_{ji}]$, the matrix $\mathcal{G}$ to be the matrix $\mathcal{G} := I - \frac{1}{2N} D + \frac{g+g^T}{2N}$ and the real number $\lambda_2(\mathcal{G})$ to be the second largest eigenvalue of a matrix $\mathcal{G}$. Let $A(t)$ to be the vector whose components are the states at time t of the N nodes. Define $A_{ave} = \sum_{i=1}^{N} \frac{A(0)}{N}$. After a number of communication rounds $N \geq N_{critical} = \frac{3 \log \epsilon^{-1}}{\log \lambda_2(\mathcal{G})}$, the gossip algorithm is guaranteed to satisfy $Pr\left(\frac{\|A(t) - A_{ave}\boldsymbol{1}\|}{\|A(0)\|} \geq \epsilon\right) \leq \epsilon$, i.e., the gossip algorithm is guaranteed ϵ-convergence with a probability of at least $1 - \epsilon$.*

The proof of the lemma is provided in [36]. We are interested in reaching equitable partitions. In order to do that we look at the problem of reaching equitable partitions as a consensus problem where the states are the areas of the regions. We use the gossip algorithm proposed as a solution to the consensus problem to solve the equitability problem.

Definition 4.3.4 (Equitably-splitting map) *A map $\mathcal{M}_e : \mathcal{P} \times \mathcal{P} \to \mathcal{P} \times \mathcal{P}$ is* equitably-splitting *if $\mathcal{M}_e(W_i, W_j) = \{W_i', W_j'\}$, implies that the following hold:*

1. $W_i' \cup W_j' = W_i \cup W_j$,

2. $\lambda(W_i' \cap W_j') = 0$,

3. $\lambda(W_i') = \lambda(W_j')$.

Such a map can be achieved using a bisection as will be shown shortly. Now that we presented the necessary maps and algorithms, we present the gossip algorithm to solve the equitable partitioning problem.

Algorithm 4.3.5 (Gossip algorithm for equitable partitioning) *We adopt the gossip algorithm presented earlier to solve the equitable partitioning problem.*

Equitable partitioning: Distributed gossip algorithm

Network: *Regions* $\{W_1, \ldots, W_N\}$ *with dual communication graph* $G = (V, E)$*, where* $(i, j) \in E \iff W_i \cap W_j \neq \phi$

State i *is* $w_i := \{W_i\}$

1: ***for*** $k = 1 : N_{critical}$

2: *choose randomly* $i, j \in \{1, \ldots, R\}$ *s.t.* $(i, j) \in E$

3: *calculate* $(W_i', W_j') = \mathcal{M}_e(W_i, W_j)$

4: $W_i := W_i'$ *and* $W_j := W_j'$;

5: ***Return:*** $\{W_1, \ldots, W_N\}$

Lemma 4.3.6 (Properties of gossip algorithms for equitable partitioning) *The properties of the gossip algorithm when applied to solve the equitable partitioning problem are the same as the properties presented in Lemma 4.3.3 of the gossip algorithm* *when applied to solve the consensus problem.*

We define the *equitably splitting* map below. As a function of the intercept $c \in \mathbb{R}$ and the angle $\theta \in [0, \pi[$, define the half-plane $H(c, \theta)$ by

$$H(c,\theta) := \{[x\ y]^T \in \mathbb{R}^2 \mid \cos(\theta)x + \sin(\theta)y + c \geq 0\}. \tag{4.2}$$

We are now ready to introduce the notion of equitably-splitting line maps.

Definition 4.3.7 (Equitably-splitting line map) *A map $\mathcal{M}_h : \mathcal{P} \times \mathcal{P} \to \mathcal{P} \times \mathcal{P}$ is an* equitably splitting line map *if the exists an angle θ, and a corresponding intercept $c(\theta) \in \mathbb{R}$, such that the following properties hold. Denoting $\mathcal{M}_h(W_i, W_j) = \{W_i', W_j')\}$, we have*

1. $\lambda(W_i' \cap W_j') = 0$,
2. $W_i'(\theta) = \{W_i \cup W_j\} \cap H(c(\theta), \theta)$ *and* $W_j'(\theta) = \{W_i \cup W_j\} \setminus W_i'(\theta)$,
3. $\lambda(W_i'(\theta)) = \lambda(W_j'(\theta))$.

Given the angle θ, a bisection algorithm can be employed to calculate the intercept $c(\theta)$ leading to an equitable splitting.

Using the separability property presented in Definition 4.1.1 of the optimization function, we solve the optimization problems by solving a sequence of pairwise optimization problems. Each pairwise optimization consists of optimally choosing W_i and W_j to maximize

$$f(W_i) + f(W_j),$$

s.t. $\lambda(W_i) = \lambda(W_j)$. The pairwise optimization problem to be solved is chosen via a *gossip* algorithm. We write the constrained pairwise optimization problem as an optimization problem over two variables. The first variable θ is the actual optimization parameter, while the second variable $c(\theta)$, is used to satisfy the optimization constraint.

We present below a family of optimization problems that we call the *separable objective and equitably constrained problem.*

Definition 4.3.8 ***(Separable objective function and equitably constrained problem)*** *Given an environment C and an objective function $f(W_1, \dots, W_N) = \sum_{i=1}^{N} g_i(W_i)$, choose a partition $\{W_1, \dots, W_N\}$ of C that maximizes*

$$f(W_1, \dots, W_N),$$

subject to $\lambda(W_i) = \lambda(W_j)$ for $i, j \in \{1, \dots, N\}$.

Lemma 4.3.9 *The problems presented in Definitions 4.2.4, 4.2.5, 4.2.8 and 4.2.9 fall under the family of problems presented in Definition 4.3.8. In addition under the assumption that the environment C is available to all nodes, the* doubly equitable problem *presented in Definition 4.2.3, is also a member of the* separable objective and equitably constrained *family of optimization problems.*

Proof: It is trivial that the problems presented in Definitions 4.2.4, 4.2.5, 4.2.8 and 4.2.9 belong to the family of the unifying problem presented in Definition 4.3.8. The only non trivial problem is the doubly equitable problem presented in Definition 4.2.3.

In fact in this case, we add to the doubly equitable partitioning problem, the assumption that the environment C is available to all regions. In this case we can write the problem in the form in Definition 4.3.8. Since $\{W_1, \dots, W_N\}$ is a partition of C, then for any measure γ, the following is true:

$$\gamma(C) = \gamma(\cup_{i=1}^N W_i) = \sum_{i=1}^N \gamma(W_i) = K.$$

And since in the doubly equitable problem, we are looking for partitions that set

$$\gamma(W_i) = \gamma(W_j)$$

for all $i, j \in \{1, \dots, N\}$, the the problem is equivalent to finding partitions $\{W_1, \dots, W_N\}$ such that

$$\gamma(W_i) = \gamma(W_j) = \frac{\gamma(C)}{N} = \frac{K}{N}.$$

So we can writing the problem as follows: Choose a partition $\{W_1, \dots, W_N\}$ of C that minimizes

$$\sum_{i=1}^N \left(\gamma(W_i) - \frac{K}{N}\right)^2,$$

subject to $\lambda(W_i) = \lambda(W_j)$ for $i, j \in \{1, 2, \dots, N\}$. This problem is obviously the same as maximizing

$$-\sum_{i=1}^N \left(\gamma(W_i) - \frac{K}{N}\right)^2,$$

subject to $\lambda(W_i) = \lambda(W_j)$ for $i, j \in \{1, 2, \dots, N\}$. This concludes the proof that the doubly equitable partitioning problem belongs to the family of problems in Definition 4.3.8.

The *bisection gossip algorithm* presented below is used to solve the family of optimization problem presented in Definition 4.3.8 where the objective function satisfies the separability property and the constraint is an equitability constraint.

Algorithm 4.3.10 (Bisection gossip algorithms) *In what follows, we restrict the partitions to be generated by half-planes. We denote by θ_{grid} the set of values that the search angle θ can take*

$$\theta_{grid} = \{\theta \in [0, \pi] \mid \theta = \frac{2\pi k}{\theta_{max}}, \text{ where } k \in \{0, \dots, \theta_{max}\}\}.$$

The distributed bisection gossip algorithm *is presented as follows.*

Bisection Gossip Algorithm # 1 **Equitable and separable optimization problems:**

Distributed bisection gossip algorithm

Network: Regions $\{W_1, \dots, W_N\}$ with dual communication graph $G = (V, E)$, where $(i, j) \in E \iff W_i \cap W_j \neq \phi$

Initialization: equitable partitions, possibly as computed by Algorithm 4.3.5

At each communication round

1: choose randomly $i, j \in \{1, \dots, R\}$ such that $(i, j) \in E$

2: compute $\mathcal{R}_{ij}(\theta_0) = f(W_i) + f(W_j) \in \mathbb{R}$

3: set $W_i^+(\theta_0) := W_i$ and $W_j^+(\theta_0) := W_j$

{ θ_0 is fictitious angle corresponding to no change in the partition. }

4: **For all** $\theta \in \theta_{grid}$

    $(W_i^+(\theta), W_j^+(\theta)) := \mathcal{M}_h(W_i, W_j)$

    compute $\mathcal{R}_{ij}(\theta) := f(W_i^+(\theta)) + f(W_j^+(\theta))$

$\theta^* = \arg\max_{\theta \in \{\theta_0 \cup \theta_{grid}\}} \mathcal{R}_{ij}(\theta)$

$W_i^+ := W_i^+(\theta^*)$ and $W_j^+ := W_j^+(\theta^*)$

For doubly equitable problems where the information about the whole environment is not available, that is when the only information available to a node is about the neighboring region with which it is communicating, we use the following algorithm.

Bisection Gossip Algorithm # 2 **Doubly equitable problems:** *Distributed bisection gossip algorithm*

Network: Regions $\{W_1, \ldots, W_N\}$ with dual communication graph $G = (V, E)$, where $(i, j) \in E \iff W_i \cap W_j \neq \phi$

Calculate equitable partitions with respect to λ as computed by Algorithm 4.3.5

At each communication round

choose randomly $i, j \in \{1, \ldots, N\}$ such that $(i, j) \in E$

compute $\mathcal{R}_{ij}(\theta_0) = (\gamma(W_i) - \gamma(W_j))^2 \in \mathbb{R}^+$

set $W_i^+(\theta_0) := W_i$ and $W_j^+(\theta_0) := W_j$

{ θ_0 is fictitious angle corresponding to no change in the partition. }

For all $\theta \in \theta_{grid}$

    $(W_i^+(\theta), W_j^+(\theta)) := \mathcal{M}_h(W_i, W_j)$

6: compute $\mathcal{R}_{ij}(\theta) := (\gamma(W_i^+(\theta)) - \gamma(W_j^+(\theta)))^2$

7: $\theta^* = \arg\min_{\theta \in \{\theta_0 \cup \theta_{grid}\}} \mathcal{R}_{ij}(\theta)$

8: $W_i^+ := W_i^+(\theta^*)$ and $W_j^+ := W_j^+(\theta^*)$

Before we present a property describing the behavior of our *bisection gossip algorithm*, we assume that the environment C is discretized.

Remark 4.3.11 (Discrete environments) *The environment C is discretized by a fixed grid. The number of points of the grid is chosen to be $grid = kR$, where $k \in \mathbb{N}$ and N is the number of nodes in the environment.*

Proposition 4.3.12 (Properties of the bisection gossip algorithm #1) Γ_{line} *is the set of fixed points of the map obtained by the* bisection gossip algorithm *map applied on a discretized environment C. The reward function is evaluated on the grid points. Along each evolution of the algorithm, the reward function is monotonically non-decreasing, and the algorithm's solution approaches the set Γ_{line}.*

Proof: In a discretized environment C, the sets of equitable partitions is finite. As the algorithm evolves, new states are chosen from a finite set of partitions. Since the algorithm only allows increases in the reward function, the reward function is monotonically non-decreasing and since the partitions only stop changing when no improvement on the reward function is possible among the *accessible finite* set of partitions, it follows that the algorithm stops changing only when Γ_{line} is reached.

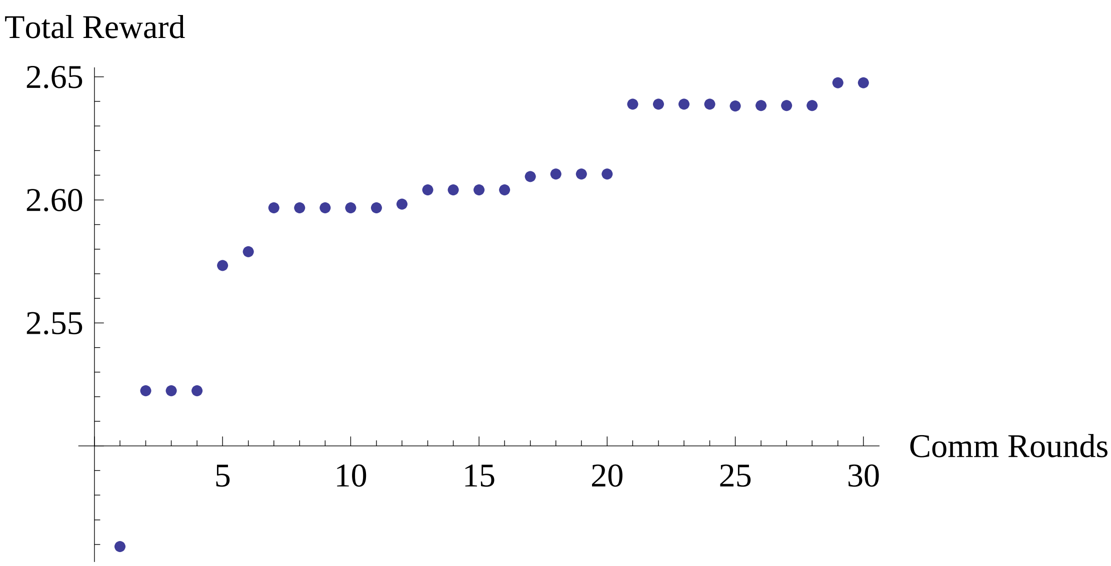


Figure 4.1: This figure shows a sample of the behavior of the reward function as the partitions are updated for each communication round when the *bisection gossip algorithm #1* is applied to a problem from the family of optimization problems presented in Definition 4.3.8.

Proposition 4.3.13 (Properties of the bisection gossip algorithm #2) *Define* Γ_{line} *to be the set of fixed points of the map obtained by the* bisection gossip algorithm *map applied on a discretized environment* C*. The reward function is evaluated on the grid points. Along each evolution of the algorithm, the algorithm's solution approaches the set* Γ_{line}.

The proof of this proposition is identical to the one presented in Proposition 4.3.12, the only difference is that in this case no claim is made about the monotonicity and of the reward function and its property of being non-decreasing.

4.4 Simulations of various environment partitioning problems

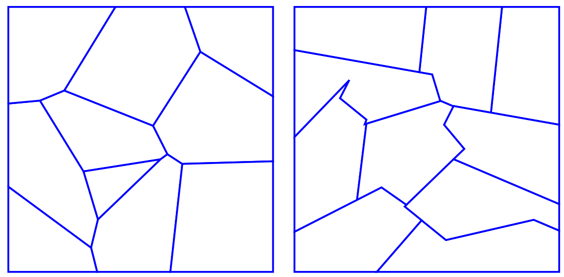
In this section we show simulation results illustrating the application of the bisection gossip algorithm to various environment partitioning problems. We will first show how our algorithm starting from a random partitioning of the environment can achieve an equitable partitioning. Unless otherwise noted, we will assume that the density over which the measure of the regions is calculated is uniform. In most of our simulations we assume uniform density, and the Lebesgue measure (except when noted otherwise).

In Figure 4.1 we show the non-decreasing monotonic property of the value of the reward function for a series of communication rounds. The algorithm applied is the *bisection gossip algorithm #1.* In Figures 4.2a and 4.2b we show how the partitions of the environment evolves from non equitable to equitable partitions under the gossip algorithm for equitable partitioning presented in Alg 4.3.5. Notice that with no further constraint than equitability of the partitions, we might obtain regions that are not convex. In Figure 4.3b we show the resulting partitions starting from the partitions in Figure 4.3a when the *bisection gossip algorithm #2* is used to solved the doubly equitable problem. In these figures the measures are the Lebesgue measure (area of the region in $2D$) and the Hausdorff measure (the perimeter of the region in $2D$). While the final value of the areas is pre-determined by the problem ($|C|/N$), various doubly equitable partitions could result in various perimeters of the regions. As is clear in the figures, the resulting partitions have equal areas and equal perimeters, but the regions could have had a smaller perimeter. The doubly equitable partitioning problem associates no cost with the length of the perimeter, or the diameter of the regions, or the isoperimetric ratio of the regions. This explains the results.

In Figures 4.4a and 4.4b, we show the optimal partitions obtained by applying the bisection gossip algorithm presented in Alg 4.3.10 to the informative partitioning problem presented in Definition 4.2.9. When all the weight is given to the performance of the localization algorithm, we end up with regions that provide little information about the position of the source in the x or y direction, as shown in Figure 4.4a. If on the other hand we add a cost associated with the diameter of the region, we end up with regions such as in Figure 4.4b that are informative in limiting the uncertainty of the position of the source in both the x and y direction.

Figure 4.5b shows the regions obtained when solving the isoperimetric problem presented in Definition 4.2.9 when applying Alg 4.3.10 starting from the partitions in Figure 4.5a. Remember that the existence of a hexagonal tiling depends on the shape and size of the region C, as well as on the number of partitions chosen.

In Figure 4.6a we show the solution to the square coverage optimization problem presented in Definition 4.2.8 starting from the partitions in Figure 4.5a. This corresponds to the informative regional localization problem where $\xi = 0$. Figure 4.6b presents the solution to the minimal perimeter problem presented in Definition 4.2.5 where Alg 4.3.10 is applied to solve the problem, starting again from the partitions in Figure 4.5a.



(a) Initial non equitable partitions (b) Equitable partitions

Figure 4.2: This figure shows a network of 10 nodes starting from random partitions in Figure 4.2a resulting in equitable partitions in Figure 4.2b after applying the algorithm.

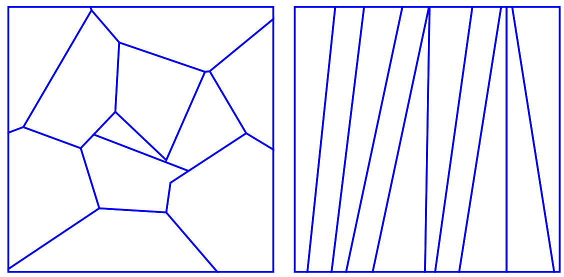
Finally, in Figure 4.7 we show the resulting partitions for the minimal perimeter equitable partitions, when the distribution is given by

$$\lambda(x,y) = \exp\{-\frac{1}{2}((x-0.9)^2 + (y-0.9)^2)\},$$

and the measure of a region W, is then given by

$$\lambda(W) = \iint\limits_{W} dx\, dy\ \lambda(x,y).$$

Note that in this figure, the regions do not have equal areas, in fact the regions around the density peak $(0.9, 0.9)$ (lightly shaded region) are smaller in area than the ones far from the density peak (darker regions).



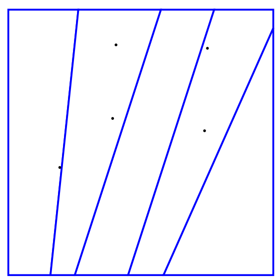
(a) Initial non equitable partitions

(b) Doubly equitable partitions

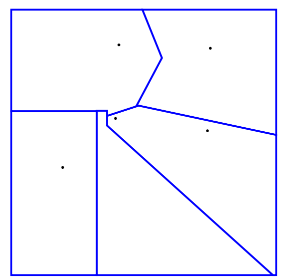
Figure 4.3: This figure shows a network of 10 nodes starting from random partitions in Figure 4.2a resulting in equitable partitions in Figure 4.2b after applying the algorithm.

4.5 Conclusion

In this work, we studied environment partitioning under various settings. In all of our problems, the variable of optimization were the partitions. We provide an algorithm that solves a class of these optimization problems as long as the term to be optimized satisfies the separability property. We implement our algorithm and find equitable partitions, optimal equitable partitions that maximize the sum of isoperimetric ratios of each partition, optimal partitions that minimize the sum of the perimeters of partitions as well as optimal partitions that minimize the sums of all the diameters of partitions and optimize general cost functions. For various density functions, various regions were obtained.

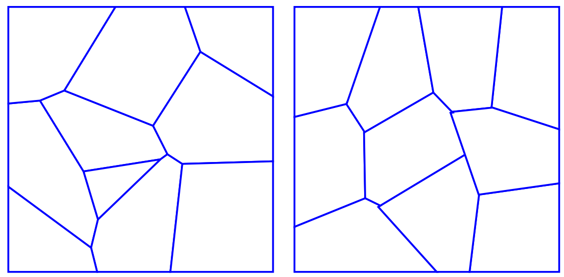


(a) Informative optimal regional partitions ξ = 1



(b) Informative optimal regional partitions ξ = 0.5

Figure 4.4: Figure 4.4a shows the partitions obtained from solving the optimization problem in Definition 4.2.9 when $\xi = 1$. Figure 4.4b shows the partitions obtained when $\xi = 0.5$.

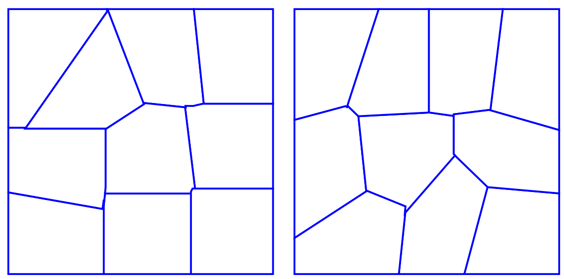


(a) Initial non equal area partitions

(b) Optimal partitions for the isoperimetric problem

Figure 4.5: This figure shows a network of 10 nodes that cooperatively partition the environment into regions that solve the isoperimetric problem presented in Definition 4.2.9.

The algorithm can be used for a wide range of density function, and we illustrate this ability of managing various densities by showing the results obtained by applying our algorithm to solve the minimal perimeter problem when the density associated with the points in the region is Gaussian. An interesting extension to this work is to allow randomization in the choice of partition, following similar methods to those presented in annealing literature, where the algorithm is allowed to take sub-optimal steps (choose sub-optimal partitions in this case) with a non zero probability. For certain problems, such an approach makes it possible to have global convergence properties that are not otherwise achievable in optimization problems of the type studied in this chapter.



(a) Informative optimal regional partitions $\xi = 0$

(b) Minimal Perimeter

Figure 4.6: Figure 4.6a shows the partitions obtained from solving the optimization problem in Definition 4.2.9 when $\xi = 0$ (or equivalently the square coverage problem). Figure 4.6b shows a network of 10 nodes that cooperatively partition the environment into regions that solve the minimal perimeter problem presented in Definition 4.2.5.

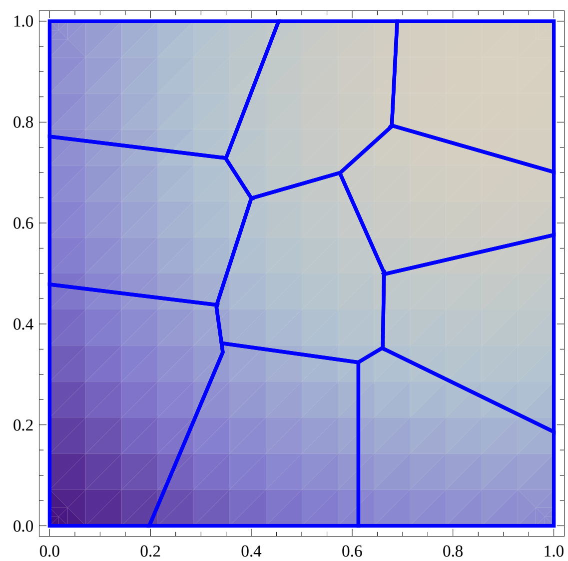


Figure 4.7: This figure shows the optimal partitions for minimal perimeter when the density function is $\lambda(x, y) = \exp\{-\frac{1}{2}((x - 0.9)^2 + (y - 0.9)^2)\}$.

Chapter 5

Conclusion and future direction

Whether our objective is to understand interactions between various individuals or simply decide on the best information processing scheme to follow, it is important to properly formulate the problem and the interaction between the final decisions reached by a group of individual; and since it is rarely, if ever, optimal to wait for every individual in the group to cast its vote or opinion before information processing starts, we found it important to study scenarios where information was sequentially processed both at the individual and the fusion center levels. With that in mind, we studied in this thesis distributed sequential decision making in groups.

5.1 Summary

In Chapter 2 we presented a framework in which we modeled sequential decision making in a group of sequential decision makers. One major difference with the problem analyzed in the literature is the sequential aspect of decision aggregation presented in this model. We were able to deal with that problem by proposing a numerical method that tracks the change in the count of decision as a function of time. We then analyzed specific rules and ran scalability analysis for large group sizes. The work presented in this chapter lays ground to many interesting questions and suggested future extensions.

In Chapter 3 we studied the problem of regional source localization where information was processed as soon as it is received. We presented the problem in a setting that made it possible to prove that algorithms can be found that have almost sure convergence. We then used sequential hypothesis testing theory to design a distributed sequential localization algorithm where each individual is able to confirm that its region contains or does not contain a source from its own information in addition to information it gathers from its neighbors. We showed that each individual satisfies a pre-defined accuracy requirements and that the test ends in a finite time.

In Chapter 4 we studied the problem of optimal regional partitioning where the setting presented in Chapter 3 guided the proposition of various optimization problems that we classified under as "separable optimization problems". We were then proposed optimization algorithms that solve problems that belong to this family.

In Chapter 5 we presented summaries of the results and future directions for the work presented here.

5.2 Open reserach directions

The problems studied in this thesis had underlying assumptions that can be relaxed to cover more general or more realistic scenarios.

Interacting sequential decision makers In the sequential decision making problem, a crucial condition for the derivation as presented was the independence of decisions between the individuals, this makes it hard to study a very interesting extension of the problem, where individuals are allowed to communicate among each other. Allowing communication between the individuals makes the problem useful in a wide range of applications ranging from social studies, to neuroscience and even systems biology, as it allows us to understand the probable steady states of a network with a group of interacting individuals, or the effects that neurons might have on one another's firing. Understanding such interactions will be useful in understanding how stochastic switching in cells will be affected in an environment where cells are competing for resources. We think that the idea of aggregating states can be further explored to understand such networks, and on the individual level interesting models where each individual can maintain its own opinion with a certain probability or might adopt the choice of the network with a complementary probability, will keep the conditional independence (conditioning here will be on the current state of the network) between the individual decisions, allowing therefore a Markov Chain model to capture the evolution of the state of the network under various scenarios. This is an on-going work.

Multiple sources localization In the regional localization problem, we studied the problem with a single source. Extending the problem to cover the multiple sources case remains an open problem. A key point needed for convergence of the localization algorithm was the uniqueness of the solution under the condition of *sufficient* data in the noise-free measurement case. In the case of multiple sources it is not clear whether such a notion of sufficiency exists, i.e., it remains an open question whether or not there is a configuration of data obtained from a number of sensors that will uniquely define the number and positions of the sources even in the noise free case. We think that pursuing this problem and then proposing localization algorithms that solve the problem in the presence of noise is a very interesting problem.

Moving sources In both the regional localization and the sequential decision making problems, we studied cases where the hypothesis are fixed. It will be interesting to account for time change in the hypothesis and analyze how the algorithms and the analysis get affected by such a change. In the regional localization problem, this corresponds to a moving source. Methods that study the problem of quickest change detection such as cumulative sums could be a good place to start in answering these questions.

Stochastic switching in groups of cells An exciting direction of some of the work presented in this thesis is the analysis of stochastic switching in cells. This extension makes great use of the scenario of interacting individuals. The problem is briefly stated here. Under stressful conditions, a group of cells take actions that preserve their DNA until the stress is removed. The behavior of the cell is a stochastic switching between two or more states, where each state has functional characteristics that improves the chances of DNA survival for the group. One caveat is that the states to which cells can switch are energy absorbing, so given the limitations on the energy available for a group, there will be a "peer aversion" like behavior by the group of cells. This aversion will be affected by the counters of the cells that switch to "energy expensive" states. We believe, that studying this problem and proposing numerical tools that allows the analysis and understanding of the behavior of the cells is of great use to the control as well as the systems biology communities and this will be part of our future directions.

Bibliography

[1] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Review of Economic Studies*, no. 14040, 2011.

[2] A. Tahbaz-Salehi, A. Sandroni, and A. Jadbabaie, "Preliminary results on learning under social influence," in *IEEE Conf. on Decision and Control*, (Shanghai, China), pp. 1513–1519, Dec. 2009.

[3] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 1945.

[4] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing* (H. V. Poor and J. B. Thomas, eds.), vol. 2, pp. 297–344, 1993.

[5] P. K. Varshney, *Distributed Detection and Data Fusion.* Signal Processing and Data Fusion, Springer, 1996.

[6] J. N. Tsitsiklis, *Problems in Decentralized Decision Making and Computation.* PhD thesis, Massachusetts Institute of Technology, Nov. 1984. Available at `http://web.mit.edu/jnt/www/Papers/PhD-84-jnt.pdf`.

[7] L. J. Young, "Computation of some exact properties of Wald's SPRT when sampling from a class of discrete distribution," *Biometrical Journal*, vol. 36, no. 5, pp. 627–637, 1994.

[8] V. V. Veeravalli, T. Başar, and H. V. Poor, "Decentralized sequential detection with sensors performing sequential tests," *Mathematics of Control, Signals and Systems*, vol. 7, no. 4, pp. 292–305, 1994.

[9] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1657–1665, 2001.

[10] W. W. Irving and J. N. Tsitsiklis, "Some properties of optimal thresholds in decentralized detection," *IEEE Transactions on Automatic Control*, vol. 39, no. 4, pp. 835–838, 1994.

[11] S. Werner and U. Noppeney, "Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization," *Cerebral Cortex*, vol. 11, pp. 1–14, 2009.

[12] S. Werner and U. Noppeney, "Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization," *Journal of Neuroscience*, vol. 30, no. 7, pp. 2662–2675, 2010.

[13] P. Laurienti, T. Perrault, T. Stanford, M. Wallace, and B. Stein, "On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies," *Experimental Brain Research*, vol. 166, pp. 289–297, 2005.

[14] D. Baranov and J. Baillieul, "Search decisions for teams of automata," in *IEEE Conf. on Decision and Control*, (Cancun, Mexico), pp. 1133–1138, 2008.

[15] M. Cao, A. Stewart, and N. Leonard, "Integrating human and robot decision-making dynamics with feedback: Models and convergence analysis," in *IEEE Conf. on Decision and Control*, (Cancun, Mexico), pp. 1127–1132, 2008.

[16] C. Woodruff, K. Morgansen, L. Vu, and D. Tomlin, "Modeling and evaluation of decision-making dynamics in sequential two-alternative forced choice tasks," in *IEEE Conf. on Decision and Control*, (Atlanta, GA), pp. 3802–3807, 2010.

[17] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: A formal analysis of performance in two-alternative forced choice tasks," *Psychological Review*, vol. 113, no. 4, pp. 700–765, 2006.

[18] N. E. Savin, "Multiple hypothesis testing," in *Handbook of Econometrics* (Z. Griliches and M. D. Intriligator, eds.), vol. 2, pp. 827–879, Elsevier, 1984.

[19] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1994–2007, 1994.

[20] P. Armitage, “Sequential analysis with more than two alternative hypotheses and its relation to discriminant function analysis,” *Journal of the Royal Statistical Society. Series B*, vol. 12, no. 1, pp. 137–144, 1950.

[21] B. Eisenberg, “Multiple hypothesis problems,” in *Handbook of Sequential Analysis* (B. K. Ghost and P. K. Sen, eds.), pp. 229–244, Mercel Dekker, 1991.

[22] T. S. Rappoport, *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[23] J. Proakis and M. Salehi, *Communication Systems Engineering*. Prentice Hall, 2001.

[24] A. H. Sayed, A. Tarighat, and Khajehnouri, “Network-based wireless location: challenges faced in developping techniques for accurate wireless location information,” *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24–40, 2005.

[25] M. Rabbat and R. Nowak, “Distributed optimization in sensor networks,” in *Symposium on Information Processing of Sensor Networks*, (Berkeley, CA), pp. 20–27, Apr. 2004.

[26] J. C. Chen, K. Yao, and R. E. Hudson, “Source localization and beamforming,” *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 30–39, 2002.

[27] M. G. Rabbat and R. D. Nowak, “Decentralized source localization and tracking,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 921–924, May 2004.

[28] G. Mao, B. Fidan, and B. D. O. Anderson, “Wireless sensor network localization techniques,” *Computer Network*, vol. 51, no. 10, pp. 2529–2553, 2007.

[29] C. Meng, Z. Ding, and S. Dasgupta, “A semidefinite programming approach to source localization in wireless network,” *IEEE Signal Processing Letters*, vol. 15, pp. 389–404, 2008.

[30] A. O. Hero III and D. Blatt, “Sensor network source localization via projection onto convex sets (PCOS),” in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 689–692, Mar. 2005.

[31] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," tech. rep., Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Nov. 2007.

[32] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, 2007. to appear.

[33] N. A. Lynch, *Distributed Algorithms*. Morgan Kaufmann, 1997.

[34] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley Series in Probability and Statistics, Wiley, 2 ed., 2000.

[35] Q. Du, V. Faber, and M. Gunzburger, "Centroidal Voronoi tessellations: Applications and algorithms," *SIAM Review*, vol. 41, no. 4, pp. 637–676, 1999.

[36] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[37] M. Pavone, A. Arsie, E. Frazzoli, and F. Bullo, "Equitable partitioning policies for robotic networks," in *IEEE Int. Conf. on Robotics and Automation*, (Kobe, Japan), pp. 2356–2361, May 2009.

[38] M. Pavone, A. Arsie, E. Frazzoli, and F. Bullo, "Distributed algorithms for environment partitioning in mobile robotic networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, 2011. To appear.

[39] J. G. Carlsson, B. Armbruster, and Y. Ye, "Finding equitable convex partitions of points in a polygon efficiently," *ACM Transactions on Algorithms*, 2008. To appear.

[40] O. Baron, O. Berman, D. Krass, and Q. Wang, "The equitable location problem on the plane," *European Journal of Operational Research*, vol. 183, no. 2, pp. 578–590, 2007.

[41] T. C. Shermer, "Recent results in art galleries," *Proceedings of the IEEE*, vol. 80, no. 9, pp. 1384–1399, 1992.

[42] V. Blasjo, "The isoperimetric problem," *Mathematical Association of America*, vol. 112, no. 6, pp. 526–566, 2005.

[43] T. Hales, "The honeycomb conjecture," *Discrete and Computation Geometry*, vol. 25, pp. 1–22, 2001.

[44] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.

[45] F. Cook, J. Ciorciari, T. Varker, and G. J. Devilly, "Changes in long term neural connectivity following psychological trauma," *Clinical Neurophysiology*, vol. 120, no. 2, pp. 309–314, 2009.

[46] Y. Liu and D. Blostein, "Optimality of the sequential probability ratio test for nonstationary observations," *IEEE Transactions on Information Theory*, vol. 38, no. 1, pp. 177–182, 1992.

[47] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2538–2541, 1999.

[48] H. Okamoto, Y. Isomura, M. Takada, and T. Fukai, "Temporal integration by stochastic recurrent network dynamics with bimodal neurons," *Journal of Neurophysiology*, vol. 97, pp. 3859–3867, 2007.

[49] P. Bauer, "A sequential elimination procedure for choosing the best population(s) based on multiple testing," *Journal of Statistical Planning and Inference*, vol. 21, no. 2, pp. 245–252, 1989.

[50] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*. Springer, 2 ed., 2000.

[51] S. H. Dandach and F. Bullo, "Algorithms for regional source localization," in *American Control Conference*, (St. Louis, MO, USA), pp. 5440–5445, June 2009.

[52] F. Bullo, R. Carli, and P. Frasca, "Gossip coverage control for robotic networks: Dynamical systems on the the space of partitions," Dec. 2009. Available at `http://arxiv.org/abs/0903.3642`.

[53] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Academic Press, 1994.

Appendix A

Proof of Arc length property

In the interest of completeness, we include this proof. We plan not to include this proof in the final version of the manuscript.

In keeping with Remark 3.2.5, consider the situation where

$$\mathbb{p}(P_i|y \in W_j)\mathbb{P}(y \in W_j) = \frac{1}{A}\int_{W_j} \delta_{\mathrm{D}}\left(\ln P_i - \ln \frac{Pd_0}{d_0 + \|q_i - y\|^\rho}\right) dy.$$

Let $q_i = [q_{i1}, q_{i2}]$ and $q_2 = [q_{21}, q_{22}]$.

Define $f : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times Q \to \mathbb{R}$ to be

$$f(q_{21}, q_{22}, P_i, q_i) = \ln P_i - \ln Pd_0 + \ln(d_0 + ((q_{i1} - q_{21})^2 + (q_{i2} - q_{22})^2)^{\rho/2}),$$

then

$$\mathbb{p}(\ln P_i|H_j)\mathbb{P}(H_j) = \frac{1}{A}\int_{W_j} \delta_{\mathrm{D}}(f(q_{21}, q_{22}, P_i, q_i))\partial q_{22}\partial q_{21}.$$

Let

$$H(a, W_j) := \{y_2 \in \mathbb{R} \mid \text{given } a \in \mathbb{R}, [a, y_2] \in W_j\}$$

Define $h : \mathbb{R} \times \mathbb{R} \times Q \to \mathbb{R}$ to be

$$\begin{aligned} h(a, P_i, q_i) &= \int_{H(a,W_j)} \delta_{\mathrm{D}}\left(\ln P_i - \ln \frac{Pd_0}{d_0 + \|q_i - y\|^\rho}\right) \partial q_{22} \\ &= \int_{H(a,W_j)} \delta_{\mathrm{D}}\left(f(a, q_{22}, P_i, q_i)\right) \partial q_{22}. \quad \text{(A.1)} \end{aligned}$$

Since

$$\frac{\partial}{\partial q_{22}} f(q_{21}, q_{22}, P_i, q_i) = f'(q_{21}, q_{22}, P_i, q_i) = \frac{\rho}{2} \cdot 2 \cdot (-1) \cdot (q_{i_2} - y_2) \frac{((q_{i1} - q_{21})^2 + (q_{i2} - q_{22})^2)^{\frac{\rho}{2}-1}}{d_0 + ((q_{i1} - q_{21})^2 + (q_{i2} - q_{22})^2)^{\frac{\rho}{2}}}$$

If we fix $q_{21} = a$, we can solve for $q_{22}(a)$ such that,

$$f(a, q_{22}(a), P_i, q_i) = 0.$$

In fact

$$f(a, q_{22}(a), P_i, q_i) = 0 \Leftrightarrow \ln P_i - \ln \frac{P d_0}{d_0 + ((q_{i1} - a)^2 + (q_{i2} - q_{22}(a))^2)^{\frac{\rho}{2}}} = 0$$

$$\Leftrightarrow (q_{i1} - a)^2 + (q_{i2} - q_{22}(a))^2 = \left(\frac{P - P_i}{P_i} d_0\right)^{\frac{2}{\rho}} = r_i^2, \quad \text{(A.2)}$$

where $r_i = \left((\frac{P}{P_i} - 1)d_0\right)^{\frac{1}{\rho}}$. Observe $H(a, W_j)$ has at most two elements satisfying equation (A.2)

$$q_{22}^1(a) = q_{i2} - \sqrt{r_i^2 - (q_{i1} - a)^2} \quad \text{(A.3)}$$

or,

$$q_{22}^2(a) = q_{i2} + \sqrt{r_i^2 - (q_{i1} - a)^2}, \quad \text{(A.4)}$$

whenever $r_i^2 \geq (q_{i1} - a)^2$. Using properties of the dirac delta function, and substituting with $q_{22}^1(a)$ and $q_{22}^2(a)$ obtained in (A.3) and (A.4), (A.1) becomes:

$$h(a, P_i, q_i) = \int_{H(a,j)} \delta_{\mathrm{D}}(f(a, q_{22}, P_i), q_i) \partial q_{22},$$

takes the values

$$h(a, P_i, q_i) = \begin{cases} \int_{H(a,W_j)} \frac{\delta_{\mathrm{D}}(q_{22} - q_{22}^1(a))}{|f'(a, q_{22}, P_i, q_i)|} \partial q_{22} \\ \text{if } q_{22}^1(a) \in H(a, W_j) \text{ but } q_{22}^2(a) \notin H(a, W_j) \\ \int_{H(a,W_j)} \frac{\delta_{\mathrm{D}}(y - q_{22}^2(a))}{|f'(a, q_{22}, P_i, q_i)|} \partial q_{22} \\ \text{if } q_{22}^2(a) \in H(a, W_j) \text{ but } q_{22}^1(a) \notin H(a, W_j) \\ \int_{H(a,W_j)} \frac{\delta_{\mathrm{D}}(q_{22} - q_{22}^1(a))}{|f'(a, q_{22}, P_i, q_i)|} + \frac{\delta_{\mathrm{D}}(q_{22} - q_{22}^2(a))}{|f'(a, q_{22}, P_i, q_i)|} \partial q_{22} \\ \text{if both } q_{22}^1(a) \text{ and } q_{22}^2(a) \in H(a, W_j) \end{cases}$$

Define $I_1(a, W_j)$, the indicator function satisfying

$$I_1(a, W_j) = \begin{cases} 1 & \text{if } q_{21}^1(a) \in H(a, W_j) \\ 0 & \text{otherwise} \end{cases}$$

Similarly define $I_2(a, W_j)$, the indicator function satisfying

$$I_2(a, W_j) = \begin{cases} 1 & \text{if } q_{22}^2(a) \in H(a, W_j) \\ 0 & \text{otherwise} \end{cases}$$

Then (A.1) becomes

$$h(a, P_i, q_i) = \frac{1}{|f'(a, q_{22}^1(a), P_i, q_i)|} I_1(a, W_j) + \frac{1}{|f'(a, q_{22}^2(a), P_i, q_i)|} I_2(a, W_j).$$

By substituting from (A.3), we get

$$\begin{aligned}\frac{1}{|f'(a, q_{22}(a), P_i, q_i)|} &= \frac{d_0 + ((q_{i1} - a)^2 + r_i^2 - (q_{i1} - a)^2)^{\frac{\rho}{2}}}{\rho\sqrt{r_i^2 - (q_{i1} - a)^2}\,((q_{i1} - a)^2 + r_i^2 - (q_{i1} - a)^2)^{\frac{\rho}{2}-1}} \\ &= \frac{d_0 + r_i^\rho}{\rho\sqrt{r_i^2 - (q_{i1} - a)^2}} \frac{1}{r_i^{\rho-2}} = \frac{d_0 + r_i^\rho}{\rho r_i^{\rho-2}} \cdot \frac{1}{\sqrt{r_i^2 - (q_{i1} - a)^2}}\end{aligned}$$

Let $\mathcal{C}_j := \{x \in \mathbb{R} \mid (q_{11}, q_{i2} - \sqrt{r_i^2 - (q_{i1} - q_{11})^2}) \in W_j\}$ and $\mathcal{C'}_j := \{x \in \mathbb{R} \mid (q_{11}, q_{i2} + \sqrt{r_i^2 - (q_{i1} - q_{11})^2}) \in W_j\}$. Note that

$$q_{11} \in \mathcal{C}_j \Rightarrow I_1(q_{11}, W_j) = 1,$$

and

$$q_{11} \in \mathcal{C'}_j \Rightarrow I_2(q_{11}, W_j) = 1.$$

Then,

$$\begin{aligned}&\mathbb{p}(P_i | y \in W_j)\mathbb{P}(y \in W_j) \\ &= \frac{1}{A}\int_{\mathcal{C}_j \bigcup \mathcal{C'}_j} h(q_{21}, P_i, q_i)\partial q_{21} = \frac{1}{A}\left(\int_{\mathcal{C}_j} h(q_{21}, P_i, q_i)\partial q_{21} + \int_{\mathcal{C'}_j} h(q_{21}, P_i, q_i)\partial q_{21}\right) \\ &= \frac{1}{A}\int_{\mathcal{C}_j} \frac{1}{|f'(q_{21}, q_{22}^1(q_{21}), P_i, q_i)|}\partial q_{21} + \frac{1}{A}\int_{\mathcal{C'}_j} \frac{1}{|f'(q_{21}, q_{22}^2(q_{21}), P_i, q_i)|}\partial q_{21}. \quad \text{(A.5)}\end{aligned}$$

Write

$$\mathcal{C}_j = \bigcup_{\alpha=1}^{s} A_\alpha \text{ , with } \bigcap_\alpha A_\alpha = \emptyset \text{ , and } A_\alpha = [a_{1_\alpha}, a_{2_\alpha}], \tag{A.6}$$

$$\mathcal{C}'_j = \bigcup_{\alpha=1}^{s'} A'_\alpha \text{ , with } \bigcap_\alpha A'_\alpha = \emptyset \text{ , and } A'_\alpha = [a'_{1_\alpha}, a'_{2_\alpha}]. \tag{A.7}$$

Equation (A.5) can then be written as the sum

$$\begin{aligned} A \cdot \mathbb{p}(P_i | y \in W_j)\mathbb{P}(y \in W_j) = &\sum_{\alpha=1}^{s} \int_{A_\alpha} \frac{1}{|f'(q_{21}, q_{22}^1(q_{21}), P_i, q_i)|} \partial q_{21} \\ &+ \sum_{\alpha=1}^{s'} \int_{A'_\alpha} \frac{1}{|f'(q_{21}, q_{22}^2(q_{21}), P_i, q_i)|} q_{21}. \end{aligned}$$

$$\begin{aligned} &A \cdot \mathbb{p}(P_i | y \in W_j)\mathbb{P}(y \in W_j) \\ &= \sum_{\alpha=1}^{s} \int_{a_{2_\alpha}}^{a_{1_\alpha}} \frac{1}{|f'(q_{21}, q_{22}^1(q_{21}), P_i, q_i)|} \partial q_{21} + \sum_{\alpha=1}^{s'} \int_{a'_{1_\alpha}}^{a'_{2_\alpha}} \frac{1}{|f'(q_{21}, q_{22}^2(q_{21}), P_i, q_i)|} \partial q_{21} \\ &\qquad = \sum_{\alpha=1}^{s} \int_{a_{2_\alpha}}^{a_{1_\alpha}} \frac{d_0 + r_i^\rho}{\rho r_i^{\rho-2}} \cdot \frac{1}{\sqrt{r_i^2 - (q_{i1} - q_{21})^2}} \partial q_{21} \\ &\qquad + \sum_{\alpha=1}^{s'} \int_{a'_{2_\alpha}}^{a'_{1_\alpha}} \frac{d_0 + r_i^\rho}{\rho r_i^{\rho-2}} \cdot \frac{1}{\sqrt{r_i^2 - (q_{i1} - q_{21})^2}} \partial q_{21} \\ &\qquad = \sum_{\alpha=1}^{s} \frac{d_0 + r_i^\rho}{\rho r_i^{\rho-2}} \cdot \arctan \frac{q_{i1} - a}{\sqrt{r_i^2 - (q_{i1} - q_{21})^2}} \Bigg|_{a_{1_\alpha}}^{a_{2_\alpha}} \\ &\qquad\qquad + \sum_{\alpha=1}^{s'} \frac{d_0 + r_i^\rho}{\rho r_i^{\rho-2}} \cdot \arctan \frac{q_{i1} - q_{21}}{\sqrt{r_i^2 - (q_{i1} - q_{21})^2}} \Bigg|_{a'_{1_\alpha}}^{a'_{2_\alpha}} \end{aligned}$$

Note that [53]

$$\begin{aligned} \arctan \frac{q_{i1} - a}{\sqrt{r_i^2 - (q_{i1} - a)^2}} &= \arctan \frac{q_{i1} - a}{\sqrt{(q_{i1} - a)^2 + (q_{i2} - q_{22}(a))^2 - (q_{i1} - a)^2}} \\ &= \arctan \frac{q_{i1} - a}{q_{i2} - q_{22}(a)} = \frac{\pi}{2} - \arctan \frac{q_{i2} - q_{22}(a)}{q_{i1} - a}. \end{aligned}$$

The conditional regional posterior becomes after simplifications,

$$\begin{aligned}
&\mathbb{p}(P_i|y \in W_j)\mathbb{P}(y \in W_j) \\
&\quad = \frac{d_0 + r_i^\rho}{A\rho r_i^{\rho-2}} \cdot \sum_{\alpha=1}^{s} \left(\frac{\pi}{2} - \arctan \frac{q_{i2} - q_{22}(a_{2_\alpha})}{q_{i1} - a_{2_\alpha}} - \frac{\pi}{2} + \arctan \frac{q_{i2} - q_{22}(a_{1_\alpha})}{q_{i1} - a_{1_\alpha}} \right) \\
&\quad + \frac{d_0 + r_i^\rho}{A\rho r_i^{\rho-2}} \sum_{\alpha=1}^{s'} \left(\frac{\pi}{2} - \arctan \frac{q_{i2} - q_{22}(a'_{2_\alpha})}{q_{i1} - a'_{2_\alpha}} - \frac{\pi}{2} + \arctan \frac{q_{i2} - q_{22}(a'_{1_\alpha})}{q_{i1} - a'_{1_\alpha}} \right) \\
&\qquad\qquad = \frac{d_0 + r_i^\rho}{A\rho r_i^{\rho-1}} \left(\sum_{\alpha=1}^{s} \text{arclength}_\alpha + \sum_{\alpha=1}^{s'} \text{arclength}'_\alpha \right),
\end{aligned}$$

where arclength_α and $\text{arclength}'_\alpha$ are the angles of the arcs in $S(W_j, r_i, q_i)$ described on distinct supports as in (A.6) and (A.7) when applicable.

Appendix B

Asymptotic and monotonicity results on combinatorial sums

Some of the results provided for the *fastest* rule and for the *majority* rule are based on the following properties of the binomial expansion $(x + y)^N = \sum_{j=0}^{N} \binom{N}{j} x^j y^{N-j}$.

Lemma B.0.1 (Properties of half binomial expansions) *For an odd number $N \in \natural$, and for real numbers $c \in \mathbb{R}$ and $x \in \mathbb{R}$ satisfying $0 < c \leq 1$ and $0 \leq x \leq c/2$, define*

$$\underline{S}(N; c, x) = \sum_{j=0}^{\lfloor N/2 \rfloor} \binom{N}{j} x^j (c - x)^{N-j}$$

and

$$\overline{S}(N; c, x) = \sum_{j=\lceil N/2 \rceil}^{N} \binom{N}{j} x^j (c - x)^{N-j}.$$

The following statements hold true:

1. *if* $0 \leq x < c/2$*, then, taking limits over odd values of* N*,*

$$\lim_{N\to\infty} \frac{\underline{S}(N;c,x)}{c^N} = 1 \qquad \textit{and} \qquad \lim_{N\to\infty} \frac{\overline{S}(N;c,x)}{c^N} = 0;$$

2. *if* $x = c/2$*, then*

$$\underline{S}(N;c,x) = \overline{S}(N;c,x) = \frac{c^N}{2};$$

3. *if* $c = 1$ *and* $0 \leq x < 1/2$*, then*

$$\overline{S}(N+2;1,x) < \overline{S}(N;1,x) \qquad \textit{and} \qquad \underline{S}(N+2;1,x) > \underline{S}(N;1,x).$$

Proof: To prove statement 1, we start with the obvious equality $c^N = (c-x+x)^N = \underline{S}(N;c,x)+\overline{S}(N;c,x)$. Therefore, it suffices to show that $\lim_{N\to\infty} \frac{\overline{S}(N;c,x)}{c^N} = 0$. Define the shorthand $h(j) := \binom{N}{j}x^j(c-x)^{N-j}$ and observe

$$\frac{h(j)}{h(j+1)} = \frac{\frac{N!}{j!(N-j)!}x^j(c-x)^{N-j}}{\frac{N!}{(j+1)!(N-j-1)!}x^{j+1}(c-x)^{N-j-1}} = \frac{j+1}{N-j}\frac{c-x}{x}.$$

It is straightforward to see that $\frac{h(j)}{h(j+1)} > 1 \iff cj - xN + c - x > 0 \iff j > \frac{xN}{c} - \frac{(c-x)}{c}$. Moreover, if $j > \frac{N}{2}$ and $0 \leq x < \frac{c}{2}$, then $j - \frac{xN}{c} + \frac{c-x}{c} > \frac{N}{2} - \frac{xN}{c} + \frac{c-x}{c} \geq \frac{N}{2} - \frac{N}{2} + \frac{c-x}{c} > 0$. Here, the second inequality follows from the fact that $-\frac{xN}{c} \geq -\frac{N}{2}$ if $0 \leq x < \frac{c}{2}$. In other words, if $j > \frac{N}{2}$ and $0 \leq x < \frac{c}{2}$, then $\frac{h(j)}{h(j+1)} > 1$. This result implies the following chain of inequalities $f(\lceil N/2\rceil) > f(\lceil N/2\rceil + 1) > \cdots > h(N)$ providing the following bound on $\overline{S}(N;c,x)$

$$\overline{S}(N;c,x) = \frac{\sum_{j=\lceil N/2\rceil}^{N} \binom{N}{j}x^j(c-x)^{N-j}}{c^N} < \frac{\lceil N/2\rceil \binom{N}{\lceil N/2\rceil}x^{\lceil N/2\rceil}(c-x)^{\lfloor N/2\rfloor}}{c^N}.$$

Since $\binom{N}{\lceil N/2\rceil} < 2^N$, we can write

$$\begin{aligned}\overline{S}(N;c,x) &< \lceil N/2\rceil \frac{2^N x^{\lceil N/2\rceil}(c-x)^{\lfloor N/2\rfloor}}{c^N} = \lceil N/2\rceil \left(\frac{2x}{c}\right)^{\lceil N/2\rceil} \left(\frac{2(c-x)}{c}\right)^{\lfloor N/2\rfloor} \\ &= \lceil N/2\rceil \left(\frac{2x}{c}\right)\left(\frac{2x}{c}\right)^{\lfloor N/2\rfloor} \left(\frac{2(c-x)}{c}\right)^{\lfloor N/2\rfloor}.\end{aligned}$$

Let $\alpha = \frac{2x}{c}$ and $\beta = 2\left(\frac{c-x}{c}\right)$ and consider $\alpha \cdot \beta = \frac{4x(c-x)}{c^2}$. One can easily show that $\alpha \cdot \beta < 1$ since $4cx - 4x^2 - c^2 = -(c-2x)^2 < 0$. The proof of statement 1 is completed by noting

$$\lim_{N\to\infty} \overline{S}(N; c, x) < \lim_{N\to\infty} \lceil N/2 \rceil \left(\frac{2x}{c}\right) (\alpha \cdot \beta)^{\lfloor N/2 \rfloor} = 0.$$

The proof of the statement 2 is straightforward. In fact it follows from the symmetry of the expressions when $x = \frac{c}{2}$, and from the obvious equality $\sum_{j=0}^{N} \binom{N}{j} x^j (c-x)^{N-j} = c^N$.

Regarding statement 3, we prove here only that $\overline{S}(N+2; 1, x) < \overline{S}(N; 1, x)$ for $0 \leq x < 1/2$. The proof of $\underline{S}(N+2; 1, x) > \underline{S}(N; 1, x)$ is analogous. Adopting the shorthand

$$f(N, x) := \sum_{i=\lceil \frac{N}{2} \rceil}^{N} \binom{N}{i} x^i (1-x)^{N-i},$$

we claim that the assumption $0 < x < 1/2$ implies

$$\Delta(N, x) := f(N+2, x) - f(N, x) < 0.$$

To establish this claim, it is useful to analyze the derivative of Δ with respect to x. We compute

$$\frac{\partial f}{\partial x}(N, x) = \sum_{i=\lceil N/2 \rceil}^{N-1} i \binom{N}{i} x^{i-1} (1-x)^{N-i} - \sum_{i=\lceil N/2 \rceil}^{N-1} (N-i) \binom{N}{i} x^i (1-x)^{N-i-1} + N x^{N-1}. \tag{B.1}$$

The first sum $\sum_{i=\lceil N/2 \rceil}^{N-1} i \binom{N}{i} x^{i-1} (1-x)^{N-i}$ in the right-hand side of (B.1) is equal to

$$\binom{N}{\lceil N/2 \rceil} \left\lceil \frac{N}{2} \right\rceil x^{\lceil N/2 \rceil - 1} (1-x)^{N - \lceil N/2 \rceil} + \sum_{i=\lceil N/2 \rceil + 1}^{N-1} i \binom{N}{i} x^{i-1} (1-x)^{N-i}.$$

Moreover, exploiting the identity $(i+1)\binom{N}{i+1} = (N-i)\binom{N}{i}$,

$$\begin{aligned}
\sum_{i=\lceil N/2 \rceil + 1}^{N-1} i \binom{N}{i} x^{i-1} (1-x)^{N-i} &= \sum_{i=\lceil N/2 \rceil}^{N-2} (i+1) \binom{N}{i+1} x^i (1-x)^{N-i-1} \\
&= \sum_{i=\lceil N/2 \rceil}^{N-2} (N-i) \binom{N}{i} x^i (1-x)^{N-i-1}.
\end{aligned}$$

The second sum in the right-hand side of (B.1) can be rewritten as

$$\sum_{i=\lceil N/2\rceil}^{N-1}(N-i)\binom{N}{i}x^i(1-x)^{N-i-1} = \sum_{i=\lceil N/2\rceil}^{N-2}(N-i)\binom{N}{i}x^i(1-x)^{N-i-1} + Nx^{N-1}.$$

Now, many terms of the two sums cancel each other out and one can easily see that

$$\begin{aligned}\frac{\partial f}{\partial x}(N,x) &= \binom{N}{\lceil N/2\rceil}\lceil N/2\rceil x^{\lceil N/2\rceil-1}\left(1-x\right)^{N-\lceil N/2\rceil}\\ &= \binom{N}{\lceil N/2\rceil}\lceil N/2\rceil\left(x\left(1-x\right)\right)^{\lceil N/2\rceil-1},\end{aligned}$$

where the last equality relies upon the identity $N-\lceil N/2\rceil = \lfloor N/2\rfloor = \lceil N/2\rceil - 1$. Similarly, we have

$$\frac{\partial f}{\partial x}(N+2,x) = \binom{N+2}{\lceil N/2\rceil+1}\left(\lceil N/2\rceil+1\right)\left(x\left(1-x\right)\right)^{\lceil N/2\rceil}.$$

Hence

$$\frac{\partial \Delta}{\partial x}(N,x) = \left(x\left(1-x\right)\right)^{\lceil N/2\rceil-1}\left(\binom{N+2}{\lceil N/2\rceil+1}\left(\lceil N/2\rceil+1\right)x(1-x) - \binom{N}{\lceil N/2\rceil}\lceil N/2\rceil\right).$$

Straightforward manipulations show that

$$\binom{N+2}{\lceil N/2\rceil+1}\left(\lceil N/2\rceil+1\right) = 4\frac{N+2}{N+1}\lceil N/2\rceil\binom{N}{\lceil N/2\rceil},$$

and, in turn,

$$\begin{aligned}\frac{\partial \Delta}{\partial x}(N,x) &= \binom{N}{\lceil N/2\rceil}\left\lceil\frac{N}{2}\right\rceil\left(x\left(1-x\right)\right)^{\lceil N/2\rceil-1}\left[4\frac{N+2}{N+1}x(1-x)-1\right]\\ &=: g(N,x)\left[4\frac{N+2}{N+1}x(1-x)-1\right],\end{aligned}$$

where the last equality defines the function $g(N,x)$. Observe that $x>0$ implies $g(N,x)>0$ and, otherwise, $x=0$ implies $g(N,x)=0$. Moreover, for all N, we have that $f(N,1/2) = 1/2$ and $f(N,0) = 0$ and in turn that $\Delta(N,1/2) = \Delta(N,0) = 0$. Additionally

$$\frac{\partial\Delta}{\partial x}(N,1/2) = g(N,1/2)\left(\frac{N+2}{N+1} - 1\right) > 0$$

and

$$\frac{\partial\Delta}{\partial x}(N,0) = 0 \qquad \text{and} \qquad \frac{\partial\Delta}{\partial x}(N,0^+) = g(N,0^+)\left(0^+ - 1\right) < 0.$$

The roots of the polynomial $x \mapsto 4\frac{N+2}{N+1}x(1-x) - 1$ are $\frac{1}{2}\left(1 \pm \sqrt{\frac{1}{N+2}}\right)$, which means that the polynomial has one root inside the interval $(0,1/2)$ and one inside the interval $(1/2,1)$. Considering all these facts together, we conclude that the function $x \mapsto \Delta(N,x)$ is strictly negative in $(0,1/2)$ and hence that $f(N+2,x) - f(N,x) < 0$.

Computation of the decision probabilities for a single SDM applying the SPRT test

In this appendix we discuss how to compute the probabilities

$$\left\{p_{\mathrm{nd}|0}\right\} \cup \left\{p_{0|0}(t), p_{1|0}(t)\right\}_{t\in\mathbb{N}} \qquad \text{and} \left\{p_{\mathrm{nd}|1}\right\} \cup \left\{p_{0|1}(t), p_{1|1}(t)\right\}_{t\in\mathbb{N}} \qquad \text{(B.2)}$$

for a single SDM applying the classical *sequential probability ratio test* (SPRT). For a short description of the SPRT test and for the relevant notation, we refer the reader to Section 2.6. We consider here observations drawn from both discrete and continuous distributions.

Appendix C

Discrete distributions of the Koopman-Darmois-Pitman form

This subsection review the procedure proposed in [7] for a certain class of discrete distributions. Specifically, [7] provides a recursive method to compute the exact values of the probabilities (B.2); the method can be applied to a broad class of discrete distributions, precisely whenever the observations are modeled as a discrete random variable of the Koopman-Darmois-Pitman form.

With the same notation as in Section 2.6, let X be a discrete random variable of the Koopman-Darmois-Pitman form; that is

$$f(x, \theta) = \begin{cases} h(x)\exp(B(\theta)Z(x) - A(\theta)), & \text{if } x \in \mathcal{Z}, \\ 0, & \text{if } x \notin \mathcal{Z}, \end{cases}$$

where $h(x)$, $Z(x)$ and $A(\theta)$ are known functions and where $\mathcal{Z}$ is a subset of the integer numbers $\mathbb{Z}$. In this section we shall assume that $Z(x) = x$. Bernoulli, binomial, geometric, negative binomial and Poisson distributions are some widely used distributions of the Koopman-Darmois-Pitman form satisfying the condition $Z(x) = x$. For distributions of this form, the likelihood associated with the t-th observation $x(t)$ is given by

$$\lambda(t) = (B(\theta_1) - B(\theta_0))x(t) - (A(\theta_1) - A(\theta_0)).$$

Let η_0, η_1 be the pre-assigned thresholds. Then, one can see that sampling will continue as long as

$$\frac{\eta_0 + t(A(\theta_1) - A(\theta_0))}{B(\theta_1) - B(\theta_0))} < \sum_{i=1}^{t} x(i) < \frac{\eta_1 + t(A(\theta_1) - A(\theta_0))}{B(\theta_1) - B(\theta_0))} \tag{C.1}$$

for $B(\theta_1) - B(\theta_0) > 0$; if $B(\theta_1) - B(\theta_0) < 0$ the inequalities would be reversed. Observe that $\sum_{i=1}^{t} x(i)$ is an integer number. Now let $\bar{\eta}_0^{(t)}$ be the smallest integer greater than $\{\eta_0 + t(A(\theta_1) - A(\theta_0))\} / (B(\theta_1) - B(\theta_0))$ and let $\bar{\eta}_1^{(t)}$ be the largest integer smaller than $\{\eta_1 + t(A(\theta_1) - A(\theta_0))\} / (B(\theta_1) - B(\theta_0))$. Sampling will continue as long as $\bar{\eta}_0^{(t)} \leq \mathcal{X}(t) \leq \bar{\eta}_1^{(t)}$ where $\mathcal{X}(t) = \sum_{i=1}^{t} x(i)$. Now suppose that, for any $\ell \in [\bar{\eta}_0^{(t)}, \bar{\eta}_1^{(t)}]$ the probability $\mathbb{P}[\mathcal{X}(t) = \ell]$ is known. Then we have

$$\mathbb{P}[\mathcal{X}(t+1) = \ell | H_i] = \sum_{j=\bar{\eta}_0^{(t)}}^{\bar{\eta}_1^{(t)}} f(\ell - j; \theta_i) \mathbb{P}[\mathcal{X}(t) = j | H_i],$$

and

$$p_{i|1}(t+1) = \sum_{j=\bar{\eta}_0^{(t)}}^{\bar{\eta}_1^{(t)}} \sum_{r=\bar{\eta}_1^{(t)}-j+1}^{\infty} \mathbb{P}[\mathcal{X}(t) = j | H_i] f(r; \theta_i),$$

$$p_{0|i}(t+1) = \sum_{j=\bar{\eta}_0^{(t)}}^{\bar{\eta}_1^{(t)}} \sum_{r=-\infty}^{\bar{\eta}_0^{(t)}-j-1} \mathbb{P}[\mathcal{X}(t) = j | H_i] f(r; \theta_i).$$

Starting with $\mathbb{P}[\mathcal{X}(0) = 1]$, it is possible to compute recursively all the quantities $\left\{p_{i|j}(t)\right\}_{t=1}^{\infty}$ and $\mathbb{P}[\mathcal{X}(t) = \ell]$, for any $t \in \mathbb{N}$, $\ell \in [\bar{\eta}_0^{(t)}, \bar{\eta}_1^{(t)}]$, and $\left\{p_{i|j}(t)\right\}_{t=1}^{\infty}$. Moreover, if the set $\mathcal{Z}$ is finite, then the number of required computations is finite.

Computation of accuracy and decision time for pre-assigned thresholds η_0 and η_1: continuous distributions

In this section we assume that X is a continuous random variable with density function given by $f(x, \theta)$. As in the previous subsection, given two pre-assigned thresholds η_0 and η_1, the goal is to compute the probabilities $p_{i|j}(t) = \mathbb{P}[\text{say} H_i | H_j, T = t]$, for $i, j \in \{1, 2\}$ and $t \in \mathbb{N}$.

We start with two definitions. Let f_{λ,θ_i} and $f_{\Lambda(t),\theta_i}$ denote, respectively, the density function of the log-likelihood function λ and of the random variable $\Lambda(t)$, under the assumption that H_i is the correct hypothesis. Assume that, for a given $t \in \mathbb{N}$, the density function $f_{\Lambda(t),\theta_i}$ is known. Then we have

$$f_{\Lambda(t),\theta_i}(s) = \int_{\eta_0}^{\eta_1} f_{\lambda,\theta_i}(s-x) f_{\Lambda(t),\theta_i}(x)dx, \qquad s \in (\eta_0, \eta_1),$$

$$p_{i|1}(t) = \int_{\eta_0}^{\eta_1} \left(\int_{\eta_1 - x}^{\infty} f_{\lambda,\theta_i}(z)dz \right) f_{\Lambda(t),\theta_i}(x)dx,$$

and

$$p_{0|i}(t) = \int_{\eta_0}^{\eta_1} \left(\int_{-\infty}^{\eta_0 - x} f_{\lambda,\theta_i}(z)dz \right) f_{\Lambda(t),\theta_i}(x)dx.$$

In what follows we propose a method to compute these quantities based on a uniform discretization of the functions λ and Λ. Interestingly, we will see how the classic SPRT algorithm can be conveniently approximated by a suitable absorbing Markov chain and how, through this approximation, the probabilities $\left\{p_{i|j}(t)\right\}_{t=1}^{\infty}$, $i, j \in \{1, 2\}$, can be efficiently computed. Next we describe our discretization approach.

First, let $\delta \in \mathbb{R}_{>0}$, $\bar{\eta}_0 = \lfloor \frac{\eta_0}{\delta} \rfloor \delta$ and $\bar{\eta}_1 = \lceil \frac{\eta_1}{\delta} \rceil \delta$. Second, for $n = \lceil \frac{\eta_1}{\delta} \rceil - \lfloor \frac{\eta_0}{\delta} \rfloor + 1$, introduce the sets

$$\mathcal{S} = \{s_1, \ldots, s_n\} \quad \text{and} \quad \Gamma = \left\{\gamma_{-n+2}, \gamma_{-n+3}, \ldots, \gamma_{-1}, \gamma_0, \gamma_1, \ldots, \gamma_{n-3}, \gamma_{n-2}\right\},$$

where $s_i = \bar{\eta}_0 + (i-1)\delta$, for $i \in \{1, \ldots, n\}$, and $\gamma_i = i\delta$, for

$$i \in \{-n+2, -n+3, \ldots, n-3, n-2\}.$$

Third, let $\bar{\lambda}$ (resp. $\bar{\Lambda}$) denote a discrete random variable (resp. a stochastic process) taking values in Γ (resp. in $\mathcal{S}$). Basically $\bar{\lambda}$ and $\bar{\Lambda}$ represent the discretization of Λ and λ, respectively. To characterize $\bar{\lambda}$, we assume that

$$\mathbb{P}\left[\bar{\lambda} = i\delta\right] = \mathbb{P}\left[i\delta - \frac{\delta}{2} \leq \lambda \leq i\delta + \frac{\delta}{2}\right], \qquad i \in \{-n+3, \ldots, n-3\},$$

$$\mathbb{P}\left[\bar{\lambda} = (-n+2)\delta\right] = \mathbb{P}\left[\lambda \leq (-n+2)\delta + \frac{\delta}{2}\right]$$

and

$$\mathbb{P}\left[\bar{\lambda} = (n-2)\delta\right] = \mathbb{P}\left[\lambda \geq (n-2)\delta - \frac{\delta}{2}\right].$$

From now on, for the sake of simplicity, we shall denote $\mathbb{P}\left[\bar{\lambda} = i\delta\right]$ by p_i. Moreover we adopt the convention that, given $s_i \in \mathcal{S}$ and $\gamma_j \in \Gamma$, we have that $s_i + \gamma_j := \bar{\eta}_0$ whenever either $i = 1$ or $i + j - 1 \leq 1$, and $s_i + \gamma_j := \bar{\eta}_1$ whenever either $i = n$ or $i + j - 1 \geq n$. In this way $s_i + \gamma_j$ is always an element of $\mathcal{S}$. Next we set $\bar{\Lambda}(t) := \sum_{h=1}^{t} \bar{\lambda}(h)$.

To describe the evolution of the stochastic process $\bar{\Lambda}$, define the row vector $\pi(t) = [\pi_1(t), \ldots, \pi_n(t)]^T \in \mathbb{R}^{1\times n}$ whose i-th component $\pi_i(t)$ is the probability that $\bar{\Lambda}$ equals s_i at time t, that is, $\pi_i(t) = \mathbb{P}\left[\bar{\Lambda}(t) = s_i\right]$. The evolution of $\pi(t)$ is described by the absorbing Markov chain $(\mathcal{S}, A, \pi(0))$ where

- $\mathcal{S}$ is the set of states with s_1 and s_n as absorbing states;
- $A = [a_{ij}]$ is the transition matrix: a_{ij} denote the probability to move from state s_i to state s_j and satisfy, according to our previous definitions and conventions,
  - $a_{11} = a_{nn} = 1$; $\quad a_{1i} = a_{nj} = 0$, for $i \in \{2, \ldots, n\}$ and $j \in \{1, \ldots, n-1\}$;
  - $a_{i1} = \sum_{s=-n+2}^{-h+1} p_s$ and $a_{in} = \sum_{s=1}^{n-2} p_s$, $\ h \in \{2, \ldots, n-1\}$;
  - $a_{ij} = p_{j-i} \quad i, j \in \{2, \ldots, n-1\}$;
- $\pi(0)$ is the initial condition and has the property that $\mathbb{P}[\bar{\Lambda}(0) = 0] = 1$.

In compact form we write $\pi(t) = \pi(0)A^t$.

The benefits of approximating the classic SPRT algorithm with an absorbing Markov chain $(\mathcal{S}, A, \pi(0))$ are summarized in the next Proposition. Before stating it, we provide some useful definitions. First, let $Q \in \mathbb{R}^{(n-2)\times(n-2)}$ be the matrix obtained by deleting the first and the last rows and columns of A. Observe that $I - Q$ is an invertible matrix and that its inverse $F := (I - Q)^{-1}$ is typically known in the literature as the *fundamental matrix* of the absorbing matrix A. Second let $A_{2:n-1}^{(1)}$ and $A_{2:n-1}^{(n)}$ denote, respectively, the first and the last column of the matrix A without the first and the last component, i.e., $A_{2:n-1}^{(1)} := \left[a_{2,1}, \ldots, a_{n-1,1}\right]^T$ and $A_{2:n-1}^{(n)} := \left[a_{2,n}, \ldots, a_{n-1,n}\right]^T$. Finally, let $e_{\lfloor \frac{\eta_0}{\delta} \rfloor + 1}$ and $\mathbf{1}_{n-2}$ denote, respectively, the vector of the canonical basis of $\mathbb{R}^{n-2}$ having 1 in the $(\lfloor \frac{\eta_0}{\delta} \rfloor + 1)$-th position and the $(n-2)$-dimensional vector having all the components equal to 1 respectively.

Proposition C.0.2 (SPRT as a Markov Chain) *Consider the classic SPRT test. Assume that we model it through the absorbing Markov chain* $(\mathcal{S}, A, \pi(0))$ *described above. Then the following statements hold:*

1. $p_{0|j}(t) = \pi_1(t) - \pi_1(t-1)$ *and* $p_{1|j}(t) = \pi_n(t) - \pi_n(t-1)$, *for* $t \in \mathbb{N}$;

2. $\mathbb{P}[\textit{say } H_0|H_j] = e^T_{\lfloor \frac{\eta_0}{\delta} \rfloor + 1} N\bar{a}_1$ *and* $\mathbb{P}[\textit{say } H_0|H_j] = e^T_{\lfloor \frac{\eta_0}{\delta} \rfloor + 1} N\bar{a}_n$; *and*

3. $\mathbb{E}[T|H_j] = e^T_{\lfloor \frac{\eta_0}{\delta} \rfloor + 1} F\mathbf{1}_{n-2}$.